

**(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION  
EN MATIÈRE DE BREVETS (PCT)**

**(19) Organisation Mondiale de la Propriété  
Intellectuelle**  
Bureau international



**(43) Date de la publication internationale**  
4 octobre 2001 (04.10.2001)

PCT

**(10) Numéro de publication internationale**  
**WO 01/72822 A2**

**(51) Classification internationale des brevets<sup>7</sup> :** C07K 14/47

**(21) Numéro de la demande internationale :** PCT/FR01/00935

**(22) Date de dépôt international :** 27 mars 2001 (27.03.2001)

**(25) Langue de dépôt :** français

**(26) Langue de publication :** français

**(30) Données relatives à la priorité :**  
00/03832 27 mars 2000 (27.03.2000) FR

**(71) Déposant (pour tous les États désignés sauf US) :** FOUNDATION JEAN DAUSSET-CEPH [FR/FR]; 27, rue Juliette Dodu, F-75010 Paris (FR).

**(72) Inventeurs; et**

**(75) Inventeurs/Déposants (pour US seulement) :** HUGOT, Jean-Pierre [FR/FR]; 5, rue de Thionville, F-75019 Paris (FR). THOMAS, Gilles [FR/FR]; 15, rue Buffon, F-75005 Paris (FR). ZOUALI, Mohamed [FR/FR]; 4, rue Bertré Albrecht, F-92220 Bagneux (FR). LESAGE, Suzanne [FR/FR]; 2, allée de la Rocade, F-78700 Conflans-Sainte-Honorine (FR). CHAMAILLARD, Mathias [FR/FR]; 3, rue des Ecureuils, F-37300 Joue-les-Tours (FR).

**(74) Mandataires :** MARTIN, Jean-Jacques etc.; Cabinet Regimbeau, 20, rue de Chazelles, F-75847 Paris Cedex 17 (FR).

**(81) États désignés (national) :** AU, CA, JP, NZ, US, ZA.

**(84) États désignés (régional) :** brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

**Publiée :**  
— sans rapport de recherche internationale, sera republiée dès réception de ce rapport

*En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.*



A2

**(54) Title: GENES INVOLVED IN INTESTINAL INFLAMMATORY DISEASES AND USE THEREOF**

**WO 01/72822** **(54) Titre : GENES IMPLIQUES DANS LES MALADIES INFLAMMATOIRES DE L'INTESTIN ET LEUR UTILISATION**

**(57) Abstract:** The invention concerns genes involved in inflammatory and/or immune diseases and some cancers, in particular intestinal cryptogenic inflammatory diseases, and proteins coded by said genes. The invention also concerns methods for diagnosing inflammatory diseases.

**(57) Abrégé :** La présente invention concerne des gènes impliqués dans les maladies inflammatoires et/ou immunes et certains cancers, en particulier les maladies inflammatoires cryptogénétiques de l'intestin, ainsi que les protéines codées par ces gènes. Des méthodes de diagnostics de maladies inflammatoires sont également des objets de la présente invention.

## GENES IMPLIQUES DANS LES MALADIES INFLAMMATOIRES DE L'INTESTIN ET LEUR UTILISATION

La présente invention concerne des gènes impliqués dans les 5 maladies inflammatoires et/ou immunes et certains cancers, en particulier les maladies inflammatoires cryptogénétiques de l'intestin, ainsi que les protéines codées par ces gènes. Des méthodes de diagnostics de maladies inflammatoires sont également des objets de la présente invention.

Les maladies inflammatoires cryptogénétiques de l'intestin (MICI) sont des 10 maladies caractérisées par une inflammation du tube digestif dont la cause est inconnue. Selon la localisation et les caractéristiques de l'inflammation on distingue deux entités nosologiques différentes: la rectocolite hémorragique (RCH) et la maladie de Crohn (MC). La RCH a été décrite par S Wilkes en 1865 tandis que le premier cas d'iléite régionale a été rapportée par Crohn en 1932. En réalité, il est 15 possible que ces deux maladies soient beaucoup plus anciennes.

Les MICI sont des maladies chroniques qui évoluent tout au long de la vie et qui touchent environ 1 à 2 personnes sur 1000 habitants dans les pays occidentaux, ce qui représente entre 60.000 et 100.000 malades en France. Il s'agit de maladies apparaissant chez le sujet jeune (le pic d'incidence est dans la troisième décennie), 20 évoluant par poussées entrecoupées de rémissions, avec des complications fréquentes telles que la dénutrition, le retard de croissance chez l'enfant, la déminéralisation osseuse et à terme la dégénérescence maligne vers le cancer du colon. Il n'existe pas de traitement spécifique. Les thérapeutiques habituelles font appel aux anti-inflammatoires, aux immunosuppresseurs et à la chirurgie. Tous ces 25 moyens thérapeutiques sont eux-mêmes source d'une morbidité iatrogène importante. Pour toutes ces raisons les MICI apparaissent comme un important problème de santé publique.

L'étiologie des MICI est actuellement inconnue. Des facteurs d'environnement sont impliqués dans la survenue de la maladie comme en 30 témoignent l'augmentation séculaire d'incidence de la maladie et la concordance incomplète chez les jumeaux monozygotes. Les seuls facteurs de risque environnementaux actuellement reconnus sont 1) le tabac dont le rôle est néfaste

dans la MC et bénéfique dans la RCH et 2) l'appendicetomie qui a un rôle protecteur pour la RCH.

Une prédisposition génétique est depuis longtemps suspectée devant l'existence d'agrégations ethniques et familiales de ces maladies. En effet, les MICI sont plus fréquentes dans la population caucasienne et en particulier la population juive d'Europe centrale. Les formes familiales représentent de 6 à 20% des cas de MICI. Elles sont particulièrement fréquentes lorsque le début de la maladie est précoce. Cependant, ce sont les études chez les jumeaux qui ont permis de confirmer le caractère génétique de ces maladies. En effet, le taux de concordance entre jumeaux pour ces maladies est plus important chez les jumeaux monozygotes que chez les jumeaux dizygotes plaident fortement pour une composante héréditaire aux MICI, en particulier à la MC. Selon toute vraisemblance, les MICI sont des maladies génétiques complexes faisant intervenir plusieurs gènes différents, en interaction entre eux et avec des facteurs d'environnement. Les MICI peuvent donc être classées dans le cadre des maladies multifactorielles.

Deux grandes stratégies ont été développées afin de mettre en évidence les gènes de susceptibilité aux MICI. La première repose sur l'analyse de gènes candidats pour des raisons physiopathologiques. Ainsi de nombreux gènes ont été proposés comme potentiellement importants pour les MICI. Il s'agit souvent de gènes ayant un rôle dans l'inflammation et la réponse immunitaire. On peut citer les gènes HLA, TAP, TNF, MICA, le récepteur T du lymphocyte, ICAM1, l'interleukine 1, CCR5, etc. D'autres gènes participent à des fonctions diverses tels que GAI2, la motilin, MRAMP, HMLH1, etc. En réalité, aucun des différents gènes candidats étudiés n'a actuellement fait la preuve définitive de son rôle dans la survenue des MICI.

Le récent développement de cartes du génome humain utilisant des marqueurs génétiques hautement polymorphes a permis aux généticiens de développer une approche non ciblée sur l'ensemble du génome. Cette démarche, appelée aussi génétique inverse ou clonage positionnel, ne fait aucune hypothèse sur les gènes impliqués dans la maladie et tente de découvrir ceux-ci à travers un criblage systématique du génome. La méthode la plus utilisée pour les maladies génétiques complexes repose sur l'étude de l'identité par la descendance des malades d'une même famille. Cette valeur est calculée pour un grand nombre (300-

400) de marqueurs de polymorphisme répartis régulièrement (tous les 10cM) sur le génome. En cas d'excès d'identité entre malades, le(s) marqueur(s) testé(s) indique(nt) une région supposée contenir un gène de susceptibilité à la maladie. Dans le cas des maladies génétiques complexes, le modèle sous-jacent à la 5 prédisposition génétique (nombre de gènes et importance respective de chacun d'entre eux) étant inconnu, les méthodes statistiques à utiliser devront être adaptées.

La présente invention concerne la mise en évidence de la séquence nucléique de gènes impliqués dans les MICI, et d'autres maladies inflammatoires, ainsi que l'utilisation de ces séquences nucléiques.

10 Dans le cadre de la présente invention, des travaux préliminaires des inventeurs ont déjà permis de localiser un gène de susceptibilité à la MC. En effet, les inventeurs (Hugot et al., 1996) ont montré qu'un gène de susceptibilité à la MC était localisé dans la région péricentromérique du chromosome 16 (figure 1). Il s'agissait du premier gène de susceptibilité à une maladie génétique complexe 15 localisé par clonage positionnel et satisfaisant aux critères stricts proposés dans la littérature (Lander et Kruglyak, 1995). Ce gène a été nommé IBD1 (pour Inflammatory Bowel Disease 1). Depuis, d'autres localisations ont été proposées par d'autres auteurs en particulier sur les chromosomes 12, 1, 3, 6 et 7 (Satsangi et al., 1996 ; Cho et al., 1998). Bien que localisés, aucun de ces gènes de susceptibilité 20 aux MICI n'a actuellement pu être identifié.

Certains auteurs n'ont pu répliquer cette localisation (Rioux et al., 1998). Ceci n'est cependant pas surprenant dans le cas de maladies génétiques complexes où une hétérogénéité génétique est probable.

Il est intéressant de noter que selon la même approche de clonage 25 positionnel, des localisations ont aussi été proposées sur le chromosome 16 pour plusieurs maladies immunes et inflammatoires telles que la spondylarthrite ankylosante, le syndrome de Blau, le psoriasis, etc. (Becker et al., 1998 ; Tromp et al., 1996). Toutes ces maladies pourraient alors partager un même gène (ou un même groupe de gènes) localisé sur le chromosome 16.

30 Le maximum des tests de liaison génétique est situé pratiquement toujours à la même position, au niveau de D16S409 ou D16S411 séparés seulement de 2cM. Ce résultat est en opposition avec la taille importante (habituellement supérieure à

20cM) de l'intervalle de confiance attribuable à la localisation génétique selon une démarche utilisant des analyses de liaison non paramétriques.

La comparaison des tests statistiques utilisés dans les travaux des inventeurs montre que les tests basés sur l'identité par descendance complète (Tz2) sont 5 meilleurs que les tests basé sur la moyenne de l'identité par descendance (Tz) (fig. 1). Une telle différence peut être expliquée par un effet récessif de IBD1.

Plusieurs gènes connus dans la région péricentromérique du chromosome 16, tels que le récepteur à l'interleukine 4, CD19, CD43, CD11, apparaissent comme de bons candidats potentiels pour la MC. Des résultats préliminaires ne plaignent 10 cependant pas en faveur de l'implication de ces gènes dans la MC.

En particulier, la présente invention fournit la séquence non seulement du gène IBD1, mais également la séquence partielle d'un autre gène, appelé IBD1prox en raison de sa localisation à proximité d'IBD, et mis en évidence comme rapporté dans les exemples ci-après. Ces gènes dont la séquence d'ADNc correspond 15 respectivement à SEQ ID N° 1 et SEQ ID N° 4 sont donc potentiellement impliqués dans de nombreuses maladies inflammatoires et/ou immunes ainsi que dans des cancers.

La séquence peptidique exprimée par les gènes IBD1 et IBD1prox est représentée par SEQ ID N° 2 et SEQ ID N° 5 respectivement; la séquence 20 génomique de ces gènes est représentée par SEQ ID N° 3 et SEQ ID N° 6 respectivement.

Ainsi, la présente invention a pour objet un acide nucléique purifié ou isolé, caractérisé en ce qu'il comprend une séquence nucléique choisie dans le groupe de séquences suivantes :

25 a) SEQ ID N° 1, SEQ ID N° 3, SEQ ID N° 4 et SEQ ID N° 6 ;  
b) la séquence d'un fragment d'au moins 15 nucléotides consécutifs d'une séquence choisie parmi SEQ ID N° 1, SEQ ID N° 3, SEQ ID N° 4 ou SEQ ID N° 6 ;  
c) une séquence nucléique présentant un pourcentage d'identité 30 d'au moins 80 %, après alignement optimal avec une séquence définie en a) ou b) ;  
d) une séquence nucléique s'hybridant dans des conditions de forte stringence avec une séquence nucléique définie en a) ou b) ;

e) la séquence complémentaire ou la séquence de l'ARN correspondant à une séquence telle que définie en a), b), c) ou d).

La séquence d'acides nucléiques selon l'invention définie en c) présente un pourcentage d'identité d'au moins 80 % après alignement optimal avec une 5 séquence telle que définie en a) ou b) ci-dessus, de préférence 90 %, de façon la plus préférée 98 %.

Par acide nucléique, séquence nucléique ou d'acide nucléique, polynucléotide, oligonucléotide, séquence de polynucléotide, séquence nucléotidique, termes qui seront employés indifféremment dans la présente 10 description, on entend désigner un enchaînement précis de nucléotides, modifiés ou non, permettant de définir un fragment ou une région d'un acide nucléique, comportant ou non des nucléotides non naturels, et pouvant correspondre aussi bien à un ADN double brin, un ADN simple brin que des produits de transcription desdits ADNs. Ainsi, les séquences nucléiques selon l'invention englobent 15 également les PNA (Peptid Nucleic Acid), ou analogues.

Il doit être compris que la présente invention ne concerne pas les séquences nucléotidiques dans leur environnement chromosomique naturel, c'est-à-dire à l'état naturel. Il s'agit de séquences qui ont été isolées et/ou purifiées, c'est-à-dire qu'elles 20 ont été prélevées directement ou indirectement, par exemple par copie, leur environnement ayant été au moins partiellement modifié. On entend ainsi également désigner les acides nucléiques obtenus par synthèse chimique.

Par « pourcentage d'identité » entre deux séquences d'acides nucléiques ou d'acides aminés au sens de la présente invention, on entend désigner un pourcentage de nucléotides ou de résidus d'acides aminés identiques entre les deux 25 séquences à comparer, obtenu après le meilleur alignement, ce pourcentage étant purement statistique et les différences entre les deux séquences étant réparties au hasard et sur toute leur longueur. On entend désigner par "meilleur alignement" ou "alignement optimal", l'alignement pour lequel le pourcentage d'identité déterminé comme ci-après est le plus élevé. Les comparaisons de séquences entre deux 30 séquences d'acides nucléiques ou d'acides aminés sont traditionnellement réalisées en comparant ces séquences après les avoir alignées de manière optimale, ladite comparaison étant réalisée par segment ou par « fenêtre de comparaison » pour identifier et comparer les régions locales de similarité de séquence. L'alignement

optimal des séquences pour la comparaison peut être réalisé, outre manuellement, au moyen de l'algorithme d'homologie locale de Smith et Waterman (1981), au moyen de l'algorithme d'homologie locale de Neddleman et Wunsch (1970), au moyen de la méthode de recherche de similarité de Pearson et Lipman (1988), au

5 moyen de logiciels informatiques utilisant ces algorithmes (GAP, BESTFIT, BLAST P, BLAST N, FASTA et TFASTA dans le Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI). Afin d'obtenir l'alignement optimal, on utilise de préférence le programme BLAST, avec la matrice BLOSUM 62. On peut également utiliser les matrices PAM ou PAM250.

10 Le pourcentage d'identité entre deux séquences d'acides nucléiques ou d'acides aminés est déterminé en comparant ces deux séquences alignées de manière optimale, la séquence d'acides nucléiques ou d'acides aminés à comparer pouvant comprendre des additions ou des délétions par rapport à la séquence de référence pour un alignement optimal entre ces deux séquences. Le pourcentage 15 d'identité est calculé en déterminant le nombre de positions identiques pour lesquelles le nucléotide ou le résidu d'acide aminé est identique entre les deux séquences, en divisant ce nombre de positions identiques par le nombre total de positions comparées et en multipliant le résultat obtenu par 100 pour obtenir le pourcentage d'identité entre ces deux séquences.

20 Par séquences nucléiques présentant un pourcentage d'identité d'au moins 80 %, de préférence 90 %, de façon plus préférée 98 %, après alignement optimal avec une séquence de référence, on entend désigner les séquences nucléiques présentant, par rapport à la séquence nucléique de référence, certaines modifications comme en particulier une délétion, une troncation, un allongement, 25 une fusion chimérique, et/ou une substitution, notamment ponctuelle, et dont la séquence nucléique présente au moins 80 %, de préférence 90 %, de façon plus préférée 98 %, d'identité après alignement optimal avec la séquence nucléique de référence. Il s'agit de préférence de séquences dont les séquences complémentaires sont susceptibles de s'hybrider spécifiquement avec les séquences SEQ ID N° 1 ou 30 SEQ ID N° 4 de l'invention. De préférence, les conditions d'hybridation spécifiques ou de forte stringence seront telles qu'elles assurent au moins 80 %, de préférence 90 %, de façon plus préférée 98 % d'identité après alignement optimal entre l'une des deux séquences et la séquence complémentaire de l'autre.

Une hybridation dans des conditions de forte stringence signifie que les conditions de température et de force ionique sont choisies de telle manière qu'elles permettent le maintien de l'hybridation entre deux fragments d'ADN complémentaires. A titre illustratif, des conditions de forte stringence de l'étape 5 d'hybridation aux fins de définir les fragments polynucléotidiques décrits ci-dessus, sont avantageusement les suivantes.

L'hybridation ADN-ADN ou ADN-ARN est réalisée en deux étapes : (1) préhybridation à 42°C pendant 3 heures en tampon phosphate (20 mM, pH 7,5) contenant 5 x SSC (1 x SSC correspond à une solution 0,15 M NaCl + 0,015 M 10 citrate de sodium), 50 % de formamide, 7 % de sodium dodécyl sulfate (SDS), 10 x Denhardt's, 5 % de dextran sulfate et 1 % d'ADN de sperme de saumon ; (2) hybridation proprement dite pendant 20 heures à une température dépendant de la taille de la sonde (i.e. : 42°C, pour une sonde de taille > 100 nucléotides) suivie de 2 lavages de 20 minutes à 20°C en 2 x SSC + 2 % SDS, 1 lavage de 20 minutes à 15 20°C en 0,1 x SSC + 0,1 % SDS. Le dernier lavage est pratiqué en 0,1 x SSC + 0,1 % SDS pendant 30 minutes à 60°C pour une sonde de taille > 100 nucléotides. Les conditions d'hybridation de forte stringence décrites ci-dessus pour un polynucléotide de taille définie, peuvent être adaptées par l'homme du métier pour des oligonucléotides de taille plus grande ou plus petite, selon l'enseignement de 20 Sambrook et al., 1989.

Parmi les séquences nucléiques présentant un pourcentage d'identité d'au moins 80 %, de préférence 90 %, de façon plus préférée 98 %, après alignement optimal avec la séquence selon l'invention, on préfère également les séquences nucléiques variantes de SEQ ID N° 1, ou de SEQ ID N° 4, ou de leurs fragments, 25 c'est-à-dire l'ensemble des séquences nucléiques correspondant à des variants alléliques, c'est-à-dire des variations individuelles des séquences SEQ ID N° 1 ou SEQ ID N° 4. Ces séquences mutées naturelles correspondent à des polymorphismes présents chez les mammifères, en particulier chez l'être humain et, notamment, à des polymorphismes pouvant conduire à la survenue d'une 30 pathologie. De préférence, la présente invention concerne les séquences nucléiques variantes dans lesquelles les mutations conduisent à une modification de la séquence d'acides aminés du polypeptide, ou de ses fragments, codés par la séquence normale de SEQ ID N° 1 ou SEQ ID N° 4.

On entend également désigner par séquence nucléique variante tout ARN ou ADNc résultant d'une mutation et/ou variation d'un site d'épissage de la séquence nucléique génomique dont l'ADNc a pour séquence SEQ ID N° 1 ou SEQ ID N° 4.

L'invention concerne de préférence un acide nucléique purifié ou isolé selon 5 la présente invention, caractérisé en ce qu'il comprend ou est constitué de l'une des séquences SEQ ID N° 1 ou SEQ ID N° 4, de leurs séquences complémentaires ou des séquences de l'ARN correspondant à SEQ ID N° 1 ou SEQ ID N° 4.

Les amores ou sondes, caractérisées en ce qu'elles comprennent une 10 séquence d'un acide nucléique selon l'invention, font également partie de l'invention.

Ainsi, la présente invention concerne également les amores ou les sondes selon l'invention qui peuvent permettre en particulier de mettre en évidence ou de discriminer les séquences nucléiques variantes, ou d'identifier la séquence génomique des gènes dont l'ADNc est représenté par SEQ ID N° 1 ou SEQ ID N° 15 4, en utilisant notamment une méthode d'amplification telle que la méthode PCR, ou une méthode apparentée.

L'invention concerne également l'utilisation d'une séquence d'acide nucléique selon l'invention comme sonde ou amorce, pour la détection, l'identification, le dosage ou l'amplification de séquence d'acide nucléique.

20 Selon l'invention, les polynucléotides pouvant être utilisés comme sonde ou comme amorce dans des procédés de détection, d'identification, de dosage ou d'amplification de séquence nucléique, présentent une taille minimale de 15 bases, de préférence de 20 bases, ou mieux de 25 à 30 bases.

Les sondes et amores selon l'invention peuvent être marquées directement 25 ou indirectement par un composé radioactif ou non radioactif par des méthodes bien connues de l'homme du métier, afin d'obtenir un signal détectable et/ou quantifiable.

Les séquences de polynucléotides selon l'invention non marquées peuvent être utilisées directement comme sonde ou amorce.

30 Les séquences sont généralement marquées pour obtenir des séquences utilisables pour de nombreuses applications. Le marquage des amores ou des sondes selon l'invention est réalisé par des éléments radioactifs ou par des molécules non radioactives.

Parmi les isotopes radioactifs utilisés, on peut citer le  $^{32}\text{P}$ , le  $^{33}\text{P}$ , le  $^{35}\text{S}$ , le  $^{3}\text{H}$  ou le  $^{125}\text{I}$ . Les entités non radioactives sont sélectionnées parmi les ligands tels la biotine, l'avidine, la streptavidine, la dioxygénine, les haptènes, les colorants, les agents luminescents tels que les agents radioluminescents, chémoluminescents, 5 bioluminescents, fluorescents, phosphorescents.

Les polynucléotides selon l'invention peuvent ainsi être utilisés comme amorce et/ou sonde dans des procédés mettant en œuvre notamment la technique de PCR (amplification en chaîne par polymérase) (Rolfs et al., 1991). Cette technique nécessite le choix de paires d'amorces oligonucléotidiques encadrant le fragment 10 qui doit être amplifié. On peut, par exemple, se référer à la technique décrite dans le brevet américain U.S. N° 4,683,202. Les fragments amplifiés peuvent être identifiés, par exemple après une électrophorèse en gel d'agarose ou de polyacrylamide, ou après une technique chromatographique comme la filtration sur gel ou la chromatographie échangeuse d'ions, puis séquencés. La spécificité de 15 l'amplification peut être contrôlée en utilisant comme amorces les séquences nucléotidiques de polynucléotides de l'invention et comme matrices, des plasmides contenant ces séquences ou encore les produits d'amplification dérivés. Les fragments nucléotidiques amplifiés peuvent être utilisés comme réactifs dans des réactions d'hybridation afin de mettre en évidence la présence, dans un échantillon 20 biologique, d'un acide nucléique cible de séquence complémentaire à celle desdits fragments nucléotidiques amplifiés.

L'invention vise également les acides nucléiques susceptibles d'être obtenus par amplification à l'aide d'amorces selon l'invention.

D'autres techniques d'amplification de l'acide nucléique cible peuvent être 25 avantageusement employées comme alternative à la PCR (PCR-like) à l'aide de couple d'amorces de séquences nucléotidiques selon l'invention. Par PCR-like on entend désigner toutes les méthodes mettant en œuvre des reproductions directes ou indirectes des séquences d'acides nucléiques, ou bien dans lesquelles les systèmes de marquage ont été amplifiés, ces techniques sont bien entendu connues. En 30 général il s'agit de l'amplification de l'ADN par une polymérase ; lorsque l'échantillon d'origine est un ARN il convient préalablement d'effectuer une transcription reverse. Il existe actuellement de très nombreux procédés permettant cette amplification, comme par exemple la technique SDA (Strand Displacement

Amplification) ou technique d'amplification à déplacement de brin (Walker et al., 1992), la technique TAS (Transcription-based Amplification System) décrite par Kwoh et al. (1989), la technique 3SR (Self-Sustained Sequence Replication) décrite par Guatelli et al. (1990), la technique NASBA (Nucleic Acid Sequence Based 5 Amplification) décrite par Kievitis et al. (1991), la technique TMA (Transcription Mediated Amplification), la technique LCR (Ligase Chain Reaction) décrite par Landegren et al. (1988), la technique de RCR (Repair Chain Reaction) décrite par Segev (1992), la technique CPR (Cycling Probe Reaction) décrite par Duck et al. (1990), la technique d'amplification à la Q-béta-réplicase décrite par Miele et al. 10 (1983). Certaines de ces techniques ont depuis été perfectionnées.

Dans le cas où le polynucléotide cible à détecter est un ARNm, on utilise avantageusement, préalablement à la mise en oeuvre d'une réaction d'amplification à l'aide des amores selon l'invention ou à la mise en œuvre d'un procédé de détection à l'aide des sondes de l'invention, une enzyme de type transcriptase 15 inverse afin d'obtenir un ADNc à partir de l'ARNm contenu dans l'échantillon biologique. L'ADNc obtenu servira alors de cible pour les amores ou les sondes mises en oeuvre dans le procédé d'amplification ou de détection selon l'invention.

La technique d'hybridation de sondes peut être réalisée de manières diverses (Matthews et al., 1988). La méthode la plus générale consiste à immobiliser l'acide 20 nucléique extrait des cellules de différents tissus ou de cellules en culture sur un support (tels que la nitrocellulose, le nylon, le polystyrène) et à incuber, dans des conditions bien définies, l'acide nucléique cible immobilisé avec la sonde. Après l'hybridation, l'excès de sonde est éliminé et les molécules hybrides formées sont détectées par la méthode appropriée (mesure de la radioactivité, de la fluorescence 25 ou de l'activité enzymatique liée à la sonde).

Selon un autre mode de mise en œuvre des sondes nucléiques selon l'invention, ces dernières peuvent être utilisées comme sondes de capture. Dans ce cas, une sonde, dite « sonde de capture », est immobilisée sur un support et sert à capturer par hybridation spécifique l'acide nucléique cible obtenu à partir de 30 l'échantillon biologique à tester et l'acide nucléique cible est ensuite détecté grâce à une seconde sonde, dite « sonde de détection », marquée par un élément facilement détectable.

Parmi les fragments d'acides nucléiques intéressants, il faut ainsi citer en particulier les oligonucléotides anti-sens, c'est-à-dire dont la structure assure, par hybridation avec la séquence cible, une inhibition de l'expression du produit correspondant. Il faut également citer les oligonucléotides sens qui, par interaction 5 avec des protéines impliquées dans la régulation de l'expression du produit correspondant, induiront soit une inhibition, soit une activation de cette expression.

Dans les deux cas (sens et anti-sens), les oligonucléotides de l'invention peuvent être utilisés *in vitro* et *in vivo*.

La présente invention concerne également un polypeptide isolé caractérisé 10 en ce qu'il comprend un polypeptide choisi parmi :

- a) un polypeptide de séquence SEQ ID N° 2 ou SEQ ID N° 5 ;
- b) un polypeptide variant d'un polypeptide de séquence définie en 15 a) ;
- c) un polypeptide homologue à un polypeptide défini en a) ou b), comportant au moins 80 % d'identité avec ledit polypeptide de a) ;
- d) un fragment d'au moins 15 acides aminés consécutifs d'un polypeptide défini en a) , b) ou c) ;
- e) un fragment biologiquement actif d'un polypeptide défini en a), 20 b) ou c).

Par « polypeptide », on entend, au sens de la présente invention, désigner des protéines ou des peptides.

Par « fragment biologiquement actif », on entend un fragment possédant la même activité biologique que le fragment peptidique dont il est déduit, de 25 préférence dans le même ordre de grandeur (à un facteur 10 près). Ainsi, les exemples montrent que la protéine IBD1 (SEQ ID N° 2) a un rôle potentiel dans les phénomènes d'apoptose. Un fragment biologiquement actif de la protéine IBD1 consiste donc en un polypeptide issu de SEQ ID N° 2 possédant également un rôle dans l'apoptose. Les exemples ci-après proposent des fonctions biologiques pour les 30 protéines IBD1 et IBD1prox, en fonction des domaines peptidiques de ces protéines et permettent ainsi à l'homme du métier d'identifier les fragments biologiquement actifs.

De préférence un polypeptide selon l'invention est un polypeptide constitué de la séquence SEQ ID N° 2 (correspondant à la protéine codée par le gène IBD1) ou de la séquence SEQ ID N° 5 (correspondant à la protéine codée par IBD1prox) ou d'une séquence possédant au moins 80 % d'identité avec SEQ ID N° 2 ou SEQ 5 ID N° 5 après alignement optimal.

La séquence du polypeptide présente un pourcentage d'identité d'au moins 80 % après alignement optimal avec les séquences SEQ ID N° 2 ou SEQ ID N° 5, de préférence 90 %, de façon plus préférée 98 %.

Par polypeptide dont la séquence d'acides aminés présentant un pourcentage 10 d'identité d'au moins 80 %, de préférence 90 %, de façon plus préférée 98 %, après alignement optimal avec une séquence de référence, on entend désigner les polypeptides présentant certaines modifications par rapport au polypeptide de référence, comme en particulier une ou plusieurs délétions, troncations, un allongement, une fusion chimérique, et/ou une ou plusieurs substitutions.

15 Parmi les polypeptides dont la séquence d'acides aminés présentant un pourcentage d'identité d'au moins 80 %, de préférence 90 %, de façon plus préférée 98 %, après alignement optimal avec les séquences SEQ ID N° 2, SEQ ID N° 5 ou avec l'un de leurs fragments selon l'invention, on préfère les polypeptides variants codés par les séquences nucléiques variantes telles que précédemment définies, en 20 particulier les polypeptides dont la séquence d'acides aminés présente au moins une mutation correspondant notamment à une troncation, délétion, substitution et/ou addition d'au moins un résidu d'acide aminé par rapport aux séquences SEQ ID N° 2, SEQ ID N° 5 ou avec l'un de leurs fragments, de manière plus préférée les polypeptides variants présentant une mutation liée à une pathologie.

25 La présente invention concerne également les vecteurs de clonage et/ou d'expression comprenant un acide nucléique ou codant pour un polypeptide selon l'invention. Un tel vecteur peut également contenir les éléments nécessaires à l'expression et éventuellement à la sécretion du polypeptide dans une cellule hôte. Une telle cellule hôte est également un objet de l'invention.

30 Les vecteurs caractérisés en ce qu'ils comportent une séquence de promoteur et/ou de régulateur selon l'invention, font également partie de l'invention.

Lesdits vecteurs comportent de préférence un promoteur, des signaux d'initiation et de terminaison de la traduction, ainsi que des régions appropriées de régulation de la transcription. Ils doivent pouvoir être maintenus de façon stable dans la cellule et peuvent éventuellement posséder des signaux particuliers 5 spécifiant la sécrétion de la protéine traduite.

Ces différents signaux de contrôle sont choisis en fonction de l'hôte cellulaire utilisé. A cet effet, les séquences d'acide nucléique selon l'invention peuvent être insérées dans des vecteurs à réplication autonome au sein de l'hôte choisi, ou des vecteurs intégratifs de l'hôte choisi.

10 Parmi les systèmes à réplication autonome, on utilise de préférence en fonction de la cellule hôte, des systèmes de type plasmidique ou viral, les vecteurs viraux pouvant notamment être des adénovirus (Perricaudet et al., 1992), des rétrovirus, des lentivirus, des poxvirus ou des virus herpétiques (Epstein et al., 1992). L'homme du métier connaît les technologies utilisables pour chacun de ces 15 systèmes.

Lorsque l'on souhaite l'intégration de la séquence dans les chromosomes de la cellule hôte, on peut utiliser par exemple des systèmes de type plasmidique ou viral ; de tels virus sont, par exemple, les rétrovirus (Temin, 1986), ou les AAV (Carter, 1993).

20 Parmi les vecteurs non viraux, on préfère les polynucléotides nus tels que l'ADN nu ou l'ARN nu selon la technique développée par la société VICAL, les chromosomes artificiels de bactérie (BAC, bacterial artificial chromosome), les chromosomes artificiels de levure (YAC, yeast artificial chromosome) pour l'expression dans la levure, les chromosomes artificiels de souris (MAC, mouse 25 artificial chromosome) pour l'expression dans les cellules murines et de manière préférée les chromosomes artificiels d'homme (HAC, human artificial chromosome) pour l'expression dans les cellules humaines.

De tels vecteurs sont préparés selon les méthodes couramment utilisées par l'homme du métier, et les clones en résultant peuvent être introduits dans un hôte 30 approprié par des méthodes standard, telles que par exemple la lipofection, l'électroporation, le choc thermique, la transformation après perméabilisation chimique de la membrane, la fusion cellulaire.

L'invention comprend en outre les cellules hôtes, notamment les cellules eucaryotes et procaryotes, transformées par les vecteurs selon l'invention ainsi que les animaux transgéniques, de préférence les mammifères, excepté l'Homme, comprenant une desdites cellules transformées selon l'invention. Ces animaux 5 peuvent être utilisés en temps que modèles, pour l'étude de l'étiologie de maladies inflammatoires et/ou immunes, et en particulier des maladies inflammatoires du tube digestif, ou pour l'étude de cancers.

Parmi les cellules utilisables aux sens de la présente invention, on peut citer les cellules bactériennes (Olins et Lee, 1993), mais aussi les cellules de levure 10 (Buckholz, 1993), de même que les cellules animales, en particulier les cultures de cellules de mammifères (Edwards et Aruffo, 1993), et notamment les cellules d'ovaire de hamster chinois (CHO). On peut citer également les cellules d'insectes dans lesquelles on peut utiliser des procédés mettant par exemple en œuvre des baculovirus (Luckow, 1993). Un hôte cellulaire préféré pour l'expression des 15 protéines de l'invention est constitué par les cellules COS.

Parmi les mammifères selon l'invention, on préfère des animaux tels que les rongeurs, en particulier les souris, les rats ou les lapins, exprimant un polypeptide selon l'invention.

Parmi les mammifères selon l'invention, on préfère également des animaux 20 tels que les souris, les rats ou les lapins, caractérisés en ce que le gène codant pour la protéine de séquence SEQ ID N° 2 ou SEQ ID N° 5, ou dont la séquence est codée par le gène homologue chez ces animaux, n'est pas fonctionnel, est invalidé ou présente au moins une mutation.

Ces animaux transgéniques sont obtenus par exemple par recombinaison 25 homologue sur cellules souches embryonnaires, transfert de ces cellules souches à des embryons, sélection des chimères affectées au niveau des lignées reproductrices, et croissance desdites chimères.

Les animaux transgéniques selon l'invention peuvent ainsi surexprimer le gène codant pour la protéine selon l'invention, ou leur gène homologue, ou 30 exprimer ledit gène dans lequel est introduite une mutation. Ces animaux transgéniques, en particulier des souris, sont obtenus par exemple par transfection de copie de ce gène sous contrôle d'un promoteur fort de nature ubiquitaire, ou sélectif d'un type de tissu, ou après transcription virale.

Alternativement, les animaux transgéniques selon l'invention peuvent être rendus déficients pour le gène codant pour l'un des polypeptides de séquences SEQ ID N° 2 ou SEQ ID N° 5, ou leurs gènes homologues, par inactivation à l'aide du système LOXP/CRE recombinase (Rohlmann et al., 1996) ou de tout autre système 5 d'inactivation de l'expression de ce gène.

Les cellules et mammifères selon l'invention sont utilisables dans une méthode de production d'un polypeptide selon l'invention, comme décrit ci-dessous, et peuvent également servir à titre de modèle d'analyse.

Les cellules ou mammifères transformés tels que décrits précédemment 10 peuvent aussi être utilisés à titre de modèles afin d'étudier les interactions entre les polypeptides selon l'invention, et les composés chimiques ou protéiques, impliqués directement ou indirectement dans les activités des polypeptides selon l'invention, ceci afin d'étudier les différents mécanismes et interactions mis en jeu.

Ils peuvent en particulier être utilisés pour la sélection de produits 15 interagissant avec les polypeptides selon l'invention, notamment la protéine de séquence SEQ ID N° 2 ou SEQ ID N° 5 ou leurs variants selon l'invention, à titre de cofacteur, ou d'inhibiteur, notamment compétitif, ou encore ayant une activité agoniste ou antagoniste de l'activité des polypeptides selon l'invention. De préférence, on utilise lesdites cellules transformées ou animaux transgéniques à titre 20 de modèle notamment pour la sélection de produits permettant de lutter contre les pathologies liées à une expression anormale de ce gène.

L'invention concerne également l'utilisation d'une cellule, d'un mammifère ou d'un polypeptide selon l'invention pour le criblage de composés chimiques ou biochimiques pouvant interagir directement ou indirectement avec les polypeptides 25 selon l'invention, et/ou capable de moduler l'expression ou l'activité de ces polypeptides.

De la même façon, l'invention concerne aussi un procédé de criblage de composés capables d'interagir *in vitro* ou *in vivo* avec un acide nucléique selon l'invention, en utilisant un acide nucléique une cellule ou un mammifère selon 30 l'invention, et en détectant la formation d'un complexe entre les composés candidats et l'acide nucléique selon l'invention.

Les composés ainsi sélectionnés sont également objets de l'invention.

L'invention concerne aussi l'utilisation d'une séquence d'acide nucléique selon l'invention pour la synthèse de polypeptides recombinants.

La méthode de production d'un polypeptide de l'invention sous forme recombinante, elle-même comprise dans la présente invention, se caractérise en ce 5 que l'on cultive les cellules transformées, notamment les cellules ou mammifères de la présente invention, dans des conditions permettant l'expression d'un polypeptide recombinant codé par une séquence d'acide nucléique selon l'invention, et que l'on récupère ledit polypeptide recombinant.

Les polypeptides recombinants, caractérisés en ce qu'ils sont susceptibles 10 d'être obtenus par ladite méthode de production, font également partie de l'invention.

Les polypeptides recombinants obtenus comme indiqué ci-dessus, peuvent aussi bien se présenter sous forme glycosylée que non glycosylée et peuvent présenter ou non la structure tertiaire naturelle.

15 Les séquences des polypeptides recombinants peuvent être également modifiées afin d'améliorer leur solubilité, en particulier dans les solvants aqueux.

De telles modifications sont connues de l'homme du métier comme par exemple la délétion de domaines hydrophobes ou la substitution d'acides aminés hydrophobes par des acides aminés hydrophiles.

20 Ces polypeptides peuvent être produits à partir des séquences d'acide nucléique définies ci-dessus, selon les techniques de production de polypeptides recombinants connues de l'homme du métier. Dans ce cas, la séquence d'acide nucléique utilisée est placée sous le contrôle de signaux permettant son expression dans un hôte cellulaire.

25 Un système efficace de production d'un polypeptide recombinant nécessite de disposer d'un vecteur et d'une cellule hôte selon l'invention.

Ces cellules peuvent être obtenues par l'introduction dans des cellules hôtes d'une séquence nucléotidique insérée dans un vecteur tel que défini ci-dessus, puis la mise en culture desdites cellules dans des conditions permettant la réplication 30 et/ou l'expression de la séquence nucléotidique transfectée.

Les procédés utilisés pour la purification d'un polypeptide recombinant sont connus de l'homme du métier. Le polypeptide recombinant peut être purifié à partir de lysats et extraits cellulaires, du surnageant du milieu de culture, par des

méthodes utilisées individuellement ou en combinaison, telles que le fractionnement, les méthodes de chromatographie, les techniques d'immunoaffinité à l'aide d'anticorps monoclonaux ou polyclonaux spécifiques, etc...

Les polypeptides selon la présente invention peuvent aussi être obtenus par 5 synthèse chimique en utilisant l'une des nombreuses synthèses peptidiques connues, par exemple les techniques mettant en œuvre des phases solides (voir notamment Stewart et al., 1984) ou des techniques utilisant des phases solides partielles, par condensation de fragments ou par une synthèse en solution classique.

Les polypeptides obtenus par synthèse chimique et pouvant comporter des 10 acides aminés non naturels correspondants sont également compris dans l'invention.

Les anticorps mono- ou polyclonaux ou leurs fragments, anticorps chimériques ou immunoconjugués, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'invention, font partie de l'invention.

15 Des anticorps polyclonaux spécifiques peuvent être obtenus à partir d'un sérum d'un animal immunisé contre les polypeptides selon l'invention, notamment produit par recombinaison génétique ou par synthèse peptidique, selon les modes opératoires usuels.

On note notamment l'intérêt d'anticorps reconnaissant de façon spécifique 20 certains polypeptides, variants, ou leurs fragments immunogènes, selon l'invention.

Les anticorps mono- ou polyclonaux ou leurs fragments, anticorps chimériques ou immunoconjugués, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement les polypeptides de séquence SEQ ID N° 2 ou SEQ ID N° 5 sont particulièrement préférés.

25 Les anticorps monoclonaux spécifiques peuvent être obtenus selon la méthode classique de culture d'hybridomes décrite par Köhler et Milstein (1975).

Les anticorps selon l'invention sont, par exemple, des anticorps chimériques, des anticorps humanisés, des fragments Fab ou F(ab')<sub>2</sub>. Ils peuvent également se présenter sous forme d'immunoconjugués ou d'anticorps marqués afin 30 d'obtenir un signal détectable et/ou quantifiable.

L'invention concerne également des méthodes pour la détection et/ou la purification d'un polypeptide selon l'invention, caractérisées en ce qu'elles mettent en œuvre un anticorps selon l'invention.

L'invention comprend en outre des polypeptides purifiés, caractérisés en ce qu'ils sont obtenus par une méthode selon l'invention.

Par ailleurs, outre leur utilisation pour la purification des polypeptides, les anticorps de l'invention, en particulier les anticorps monoclonaux, peuvent 5 également être utilisés pour la détection de ces polypeptides dans un échantillon biologique.

Ils constituent ainsi un moyen d'analyse immunocytochimique ou immuno-histochimique de l'expression des polypeptides selon l'invention, notamment les polypeptides de séquence SEQ ID N° 2, SEQ ID N° 5 ou l'un de leurs variants, sur 10 des coupes de tissus spécifiques, par exemple par immunofluorescence, marquage à l'or, immuno-conjugués enzymatiques.

Ils peuvent permettre notamment de mettre en évidence une expression anormale de ces polypeptides dans les tissus ou prélèvements biologiques.

Plus généralement, les anticorps de l'invention peuvent être 15 avantageusement mis en œuvre dans toute situation où l'expression d'un polypeptide selon l'invention, normal ou muté, doit être observée.

Ainsi, un procédé de détection d'un polypeptide selon l'invention dans un échantillon biologique, comprenant les étapes de mise en contact de l'échantillon biologique avec un anticorps selon l'invention et de mise en évidence du complexe 20 antigène-anticorps formé est également un objet de l'invention, ainsi qu'une trousse permettant de mettre en œuvre un tel procédé. Une telle trousse contient en particulier :

- a) un anticorps monoclonal ou polyclonal selon l'invention ;
- b) éventuellement des réactifs pour la constitution d'un milieu 25 propice à la réaction immunologique ;
- c) les réactifs permettant la détection du complexe antigène-anticorps produit lors de la réaction immunologique.

Les anticorps selon l'invention peuvent également être utilisés dans le traitement d'une maladie inflammatoire et/ou immune, ou d'un cancer, chez 30 l'homme, lorsque l'on observe une expression anormale du gène IBD1 ou du gène IBD1prox. Une expression anormale signifie une surexpression ou l'expression d'une protéine mutée.

Ces anticorps peuvent être obtenus directement à partir de sérum humain, ou à partir d'animaux immunisés avec des polypeptides selon l'invention, puis « humanisés », et peuvent être utilisés tels quels ou dans la préparation d'un médicament destiné au traitement des maladies précitées.

5 Font également partie de l'invention, les méthodes de détermination d'une variabilité allélique, d'une mutation, d'une délétion, d'une perte d'hétérozygotie ou de toute anomalie génétique du gène selon l'invention, caractérisées en ce qu'elles mettent en oeuvre une séquence d'acide nucléique, un polypeptide ou un anticorps selon l'invention.

10 L'invention fournit en effet la séquence des gènes IBD1 et IBD1prox impliqués dans des maladies inflammatoires et/ou immunes, et en particulier les MICI. Un des enseignements de l'invention est de préciser les mutations dans ces séquences nucléiques ou polypeptidiques, qui sont liées à un phénotype correspondant à une des ces maladies inflammatoires et/ou immunes.

15 On peut détecter ces mutations directement par analyse de l'acide nucléique et des séquences selon l'invention (ADN génomique, ARN, ou ADNc), mais également par l'intermédiaire des polypeptides selon l'invention. En particulier, l'utilisation d'un anticorps selon l'invention qui reconnaît un épitope portant une mutation permet de discriminer entre une protéine « saine » et une protéine 20 « associée à une pathologie ».

Ainsi, l'étude du gène IBD1 dans diverses maladies inflammatoires et/ou immunes humaines montre ainsi qu'il existe des variants de séquence de ce gène dans la maladie de Crohn, la rectocolite hémorragique et le syndrome de Blau, comme démontré par les exemples. Ces variations de séquence aboutissent à des 25 variations importantes de la séquence protéique déduite. En effet, elles sont soit localisées sur des sites très conservés de la protéine dans des domaines fonctionnels importants, soit elles aboutissent à la synthèse d'une protéine tronquée. Il est donc extrêmement probable que ces altérations entraînent une modification de la fonction de la protéine et aient donc un effet causal dans la survenue de ces maladies.

30 La variété des maladies où sont observées ces mutations suggère que le gène IBD1 est potentiellement important dans de nombreuses maladies inflammatoires et/ou immunes. Ce résultat est à rapprocher du fait que la région péricentromérique du chromosome 16 a été décrite comme contenant des gènes de susceptibilité à

diverses maladies humaines telles que la spondylarthrite ankylosante ou le rhumatisme psoriasique. On peut donc considérer qu'IBD1 a un rôle important dans un grand nombre de maladies inflammatoires et/ou immunes.

En particulier, on peut associer IBD1 aux maladies inflammatoires 5 granulomateuses. En effet, le Syndrome de Blau et la MC sont des maladies faisant partie de cette famille. On espère donc trouver des variations dans le gène IBD1 pour les autres maladies de la même famille (sarcoïdose, maladie de Behçet...).

De plus, l'implication de IBD1 dans les voies cellulaires aboutissant à l'apoptose soulève la question de son éventuel rôle carcinogène. En effet, il est 10 attendu qu'une dysrégulation de IBD1 puisse aboutir à une prédisposition cancéreuse. Cette hypothèse est renforcée par le fait qu'il existe une prédisposition au cancer du colon dans les maladies inflammatoires de l'intestin. IBD1 pourrait en partie expliquer cette susceptibilité au cancer et définir de nouvelles voies de carcinogenèse.

15 La description précise des mutations observables dans le gène IBD1 permet ainsi de poser les bases d'un diagnostic moléculaire des maladies inflammatoires et immunes où son rôle est démontré. Une telle démarche, basée sur la recherche de mutations dans le gène, permettra de contribuer au diagnostic de ces maladies et éventuellement de réduire l'importance de certains examens complémentaires 20 invasifs ou coûteux. L'invention pose les bases d'un tel diagnostic moléculaire basé sur la recherche de mutations dans IBD1.

Le diagnostic moléculaire des maladies inflammatoires devrait aussi permettre d'améliorer la classification nosologique de ces maladies et de mieux définir des sous-groupes de malades particuliers par leur caractéristiques cliniques, 25 l'évolutivité de la maladie ou la réponse à certains traitements. A titre d'exemple, le démembrément des mutations existantes pourrait ainsi permettre de classer les colites actuellement indéterminées qui représentent plus de 10% des maladies inflammatoires de l'intestin. Une telle démarche permettra de proposer une prise en charge précoce adaptée à chaque patient. D'une manière générale, une telle 30 démarche permet d'espérer pouvoir définir à terme une prise en charge individualisée de la maladie, en fonction du terrain génétique de chaque malade, incluant des mesures curatives et préventives.

En particulier, on préfère une méthode de diagnostic et/ou d'évaluation pronostique d'une maladie inflammatoire ou d'un cancer caractérisée en ce qu'on détermine à partir d'un prélèvement biologique d'un patient la présence d'au moins une mutation et/ou une altération d'expression du gène correspondant à SEQ ID N° 5 1 ou SEQ ID N° 4 par l'analyse de tout ou partie d'une séquence nucléique correspondant audit gène. On peut aussi étudier les gènes SEQ ID N° 3 ou SEQ ID N° 6.

Cette méthode de diagnostic et/ou d'évaluation pronostique peut être utilisée de façon préventive (étude d'une prédisposition à ces maladies inflammatoires ou 10 au cancer), ou afin de servir à l'établissement et/ou la confirmation d'un état clinique chez un patient.

De préférence, la maladie inflammatoire est une maladie inflammatoire du tube digestif, et le cancer est un cancer du tube digestif (intestin grêle ou colon).

L'enseignement de l'invention permet en effet de connaître les mutations 15 présentant un déséquilibre de liaison avec les maladies inflammatoires du tube digestif, et qui sont donc associées à de telles maladies.

L'analyse peut être effectuée par séquence de tout ou partie du gène, ou par d'autres méthodes connues de l'homme du métier. On peut en particulier utiliser 20 des méthodes basées sur la PCR, par exemple la PCR-SSCP qui permet de détecter des mutations ponctuelles.

On peut également effectuer l'analyse par fixation d'une sonde selon l'invention correspondant à l'une des séquences SEQ ID N° 1, 3, 4 ou 6 sur une puce à ADN et l'hybridation sur ces microplaques. Une puce à ADN contenant une séquence selon l'invention est également un des objets de l'invention.

25 De même, une puce à protéines contenant une séquence d'acides aminés selon l'invention est aussi un objet de l'invention. Une telle puce à protéines permet l'étude des interactions entre les polypeptides selon l'invention et d'autres protéines ou des composés chimiques, et peut ainsi être utile pour le criblage de composés interagissant avec les polypeptides selon l'invention. On peut également utiliser les 30 puces à protéines selon l'invention pour détecter la présence d'anticorps dirigés contre les polypeptides selon l'invention dans le sérum de patients. On peut aussi mettre en œuvre une puce à protéines contenant un anticorps selon l'invention.

L'homme du métier sait également mettre en œuvre des techniques permettant l'étude de l'altération de l'expression d'un gène, par exemple par l'étude de l'ARNm (en particulier par Northern Blot ou par des expériences de RT-PCR, avec des sondes ou des amorces selon l'invention), ou de la protéine exprimée, en 5 particulier par Western Blot, en utilisant des anticorps selon l'invention.

Le gène testé est de préférence le gène de séquence SEQ ID N° 1, la maladie inflammatoire pour laquelle on cherche à prédire la susceptibilité étant une maladie du tube digestif, en particulier la maladie de Crohn, ou la rectocolite hémorragique. Si l'on cherche à détecter un cancer, il s'agit de préférence du cancer du colon.

10 L'invention se rapporte également à des procédés d'obtention d'un allèle du gène IBD1, associé à un phénotype détectable, comprenant les étapes suivantes :

- a) obtenir un échantillon d'acide nucléique d'un individu exprimant ledit phénotype détectable ;
- b) mettre en contact ledit échantillon d'acide nucléique avec un 15 agent capable de détecter spécifiquement un acide nucléique codant pour la protéine IBD1 ;
- c) isoler ledit acide nucléique codant pour la protéine IBD1.

Un tel procédé peut être suivi d'une étape de séquence de tout ou partie de l'acide nucléique codant pour la protéine IBD1, ce qui permet de prédire la 20 susceptibilité à une maladie inflammatoire ou d'un cancer.

L'agent capable de détecter spécifiquement un acide nucléique codant pour la protéine IBD1 est avantageusement une sonde d'oligonucléotides selon l'invention, qui peut être formée d'ADN, d'ARN, de PNA, modifiés ou non. Les modifications peuvent inclure un marquage radioactif ou fluorescent, ou être dues à 25 des modifications dans les liaisons entre les bases (phosphorothioates, ou méthylphosphonates par exemple). L'homme du métier connaît les protocoles permettant d'isoler une séquence spécifique d'ADN. L'étape b) du procédé ci-dessus décrit peut également être une étape d'amplification telle que décrite précédemment.

30 L'invention se rapporte également à un procédé de détection et/ou de dosage d'un acide nucléique selon l'invention dans un échantillon biologique, comprenant les étapes suivantes de mise en contact d'une sonde selon l'invention avec un

échantillon biologique et de détection et/ou dosage de l'hybride formé entre ledit polynucléotide et l'acide nucléique de l'échantillon biologique.

L'homme du métier sait mettre en œuvre un tel procédé, et peut en particulier utiliser une trousse de réactifs comprenant :

- 5 a) un polynucléotide selon l'invention, utilisé en tant que sonde ;
- b) les réactifs nécessaires à la mise en œuvre d'une réaction d'hybridation entre ladite sonde et l'acide nucléique de l'échantillon biologique ;
- c) les réactifs nécessaires à la détection et/ou le dosage de l'hybride

10 formé entre ladite sonde et l'acide nucléique de l'échantillon biologique ;

qui est également un objet de l'invention.

Une telle trousse peut également contenir des contrôles positifs ou négatifs afin d'assurer la qualité des résultats obtenus.

15 Toutefois, afin de détecter et/ou doser un acide nucléique selon l'invention, l'homme du métier peut également effectuer une étape d'amplification à l'aide d'amorces choisies parmi les séquences selon l'invention.

Enfin, l'invention concerne également les composés choisis parmi un acide nucléique, un polypeptide, un vecteur, une cellule, ou un anticorps selon 20 l'invention, ou les composés obtenus par les procédés de criblage selon l'invention, à titre de médicament, en particulier pour la prévention et/ou le traitement d'une maladie inflammatoire et/ou immune ou d'un cancer, associé à la présence d'au moins une mutation du gène correspondant à SEQ ID N° 1 ou SEQ ID N° 4, de préférence une maladie inflammatoire du tube digestif, en particulier la maladie de 25 Crohn ou la rectocolite hémorragique.

Les exemples qui suivent permettent de mieux comprendre les avantages de l'invention et ne doivent pas être considérés comme limitant la portée de l'invention.

#### **DESCRIPTION DES FIGURES**

30 Figure 1 : tests de liaison génétique non paramétrique pour la maladie de Crohn dans la région péricentromérique du chromosome 16 (d'après Hugot et al., 1996). Analyse de liaison multipoint basé sur l'identité par descendance pour les marqueurs de la région péricentromérique du chromosome 16. Les distances

génétiques entre marqueurs ont été estimées grâce au programme CRIMAP. Le lod score (MAPMAKER/SIBS) est indiqué sur la figure de gauche. Deux tests de pseudo vraisemblance ont été développés et rapportés sur la figure de droite. Le premier ( $T_z$ ) est analogue au test des moyennes. Le deuxième ( $T_z2$ ) est analogue au test de la proportion des paires d'affectés partageant deux allèles.

5 Figure 2 : analyse de liaison génétique multipoint non paramétrique. 78 familles avec plusieurs apparentés atteints de Maladie de Crohn ont été génotypées pour 26 marqueurs de polymorphisme dans la région péricentromérique du chromosome 16. La localisation de chaque marqueur est symbolisée par une flèche. L'ordre des 10 marqueurs et la distance les séparant dérive de l'analyse des données expérimentales avec le logiciel Crimap. Les flèches sous la courbe indiquent les marqueurs SPN, D16S409 et D16S411 utilisés dans la première étude publiée (Hugot et al., 1996). Les flèches situées en haut de la figure correspondent aux marqueurs D16S3136, D16S541, D16S3117, D16S416 et D16S770 localisés au 15 maximum du test de liaison génétique. Les données de typage ont été analysées à l'aide du programme d'analyse multipoint non paramétrique du logiciel Genehunter version 1.3. Le maximum du NPL Score est de 3,33 ( $p=0,0004$ ).

10 Figure 3 : représentation schématique de la protéine codée par IBD1. La protéine codée par IBD1 est représentée horizontalement. Les différents domaines qui la 20 composent sont indiqués sur la figure avec le numéro de référence des acides aminés correspondant au début et à la fin de chaque domaine. La protéine est constituée d'un domaine CARD, d'un domaine liant les nucléotides (NBD) et de 25 motifs riches en leucines (LRR).

Figure 4 : représentation schématique de la protéine IBD1/NOD2 dans trois variants associés à MC.

30 A : Le produit de traduction déduit de la séquence d'ADNc du gène candidat IBD1 est identique à celui de NOD2 (Ogura et al., 2000). Le polypeptide contient 2 domaines CARD (CAspase Recruitment Domains), un domaine de liaison aux nucléotides (NBD) et 10 répétitions de 27 acides aminés, des motifs riches en 35 leucine (LRR). La séquence consensus du site du motif A (boucle P) liant l'ATP/GTP du NBD est indiquée par un cercle noir. Les changements de séquences codés par les trois principaux variants associés à MC sont SNP 8 (R675W), SNP 12 (G881R) et SNP 13 (déplacement de cadre 980). Le déplacement de cadre change

un codon leucine en un codon proline à la position 980 qui est immédiatement suivi par un codon stop.

B : Variants faux sens rares de NOD2 chez 457 patients MC, 159 patients RCH et 103 individus non apparentés, non atteints. Les positions des variants faux sens rares sont indiquées pour les trois groupes. L'échelle à gauche indique le nombre de chaque variant identifié dans les groupes faisant l'objet de recherche et celle à droite mesure la fréquence de la mutation. Les fréquences alléliques du polymorphisme V928I n'étaient pas significativement différentes (0,92 : 0,08) dans les trois groupes et les génotypes correspondants étaient en équilibre Hardy-Weinberg.

## EXEMPLES

### Exemple 1 : localisation fine de IBD1

La première étape vers l'identification du gène IBD1 a été de réduire la taille de la région génétique d'intérêt, initialement centrée sur le marqueur D16S411 situé entre D16S409 et D16S419 (Hugot et al., 1996 et fig. 1). Un groupe de marqueurs proches (carte génétique à haute résolution) a été utilisé pour mieux préciser la région génétique et a permis de compléter les analyses de liaison génétique et de rechercher un déséquilibre de liaison génétique avec la maladie.

L'étude a porté sur 78 familles comportant au moins 2 apparentés atteints de MC, qui correspondaient à 119 paires d'affectés. Les familles comportant des malades atteints de RCH ont été exclues de l'étude.

Vingt-six marqueurs génétiques de polymorphisme de type microsatellites ont été étudiés. Ces marqueurs formaient ensemble une carte à haute résolution avec une distance moyenne entre marqueurs de l'ordre de 1cM dans la région génétique d'intérêt. Les caractéristiques des marqueurs étudiés sont rapportés sur le tableau 1.

Tableau 1. Marqueurs polymorphes de type microsatellite utilisés pour la localisation fine de IBD1

| Nom du marqueur de polymorphisme | Distance cumulée (cM) | Amorces PCR                 |
|----------------------------------|-----------------------|-----------------------------|
| D16S3120<br>(AFM326vc5)          | 0                     | SEQ ID N° 7<br>SEQ ID N° 8  |
| D16S298<br>(AFMa189wg5)          | 2,9                   | SEQ ID N° 9<br>SEQ ID N° 10 |

|                          |      |                              |
|--------------------------|------|------------------------------|
| D16S299                  | 3,4  | SEQ ID N° 11<br>SEQ ID N° 12 |
| SPN                      | 3,9  | SEQ ID N° 13<br>SEQ ID N° 14 |
| D16S383                  | 4,3  | SEQ ID N° 15<br>SEQ ID N° 16 |
| D16S753<br>(GGAA3G05)    | 4,9  | SEQ ID N° 17<br>SEQ ID N° 18 |
| D16S3044<br>(AFMa222za9) | 5,8  | SEQ ID N° 19<br>SEQ ID N° 20 |
| D16S409<br>(AFM161xa1)   | 5,8  | SEQ ID N° 21<br>SEQ ID N° 22 |
| D16S3105<br>(AFMb341zc5) | 6,1  | SEQ ID N° 23<br>SEQ ID N° 24 |
| D16S261<br>(MFD24)       | 6,8  | SEQ ID N° 25<br>SEQ ID N° 26 |
| D16S540<br>(GATA7B02)    | 6,9  | SEQ ID N° 27<br>SEQ ID N° 28 |
| D16S3080<br>(AFMb068zb9) | 7    | SEQ ID N° 29<br>SEQ ID N° 30 |
| D16S517<br>(AFMa132we9)  | 7    | SEQ ID N° 31<br>SEQ ID N° 32 |
| D16S411<br>(AFM186xa3)   | 8    | SEQ ID N° 33<br>SEQ ID N° 34 |
| D16S3035<br>(AFMa189wg5) | 10,4 | SEQ ID N° 35<br>SEQ ID N° 36 |
| D16S3136<br>(AFMa061xe5) | 10,4 | SEQ ID N° 37<br>SEQ ID N° 38 |
| D16S541<br>(GATA7E02)    | 11,4 | SEQ ID N° 39<br>SEQ ID N° 40 |
| D16S3117<br>(AFM288wb1)  | 11,5 | SEQ ID N° 41<br>SEQ ID N° 42 |
| D16S416<br>(AFM210yg3)   | 12,4 | SEQ ID N° 43<br>SEQ ID N° 44 |
| D16S770<br>(GGAA20G02)   | 13,2 | SEQ ID N° 45<br>SEQ ID N° 46 |
| D16S2623<br>(GATA81B12)  | 15   | SEQ ID N° 47<br>SEQ ID N° 48 |
| D16S390                  | 16,5 | SEQ ID N° 49<br>SEQ ID N° 50 |
| D16S419<br>(AFM225zf2)   | 20,4 | SEQ ID N° 51<br>SEQ ID N° 52 |
| D16S771<br>(GGAA23C09)   | 21,8 | SEQ ID N° 53<br>SEQ ID N° 54 |
| D16S408<br>(AFM137xf8)   | 25,6 | SEQ ID N° 55<br>SEQ ID N° 56 |
| D16S508<br>(AFM304xf1)   | 38,4 | SEQ ID N° 57<br>SEQ ID N° 58 |

Chaque marqueur est répertorié selon la nomenclature internationale et le plus souvent par le nom proposé par le laboratoire d'origine. Les marqueurs apparaissent selon leur ordre sur le chromosome (de 16p vers 16q). La distance génétique entre les marqueurs (en centiMorgan Kosambi, calculée par le programme Crimap à partir des données expérimentales) est indiquée dans la deuxième colonne. Le premier marqueur polymorphe est pris arbitrairement comme point de référence. Les oligonucléotides ayant servi à la réaction de polymérisation en chaîne (PCR) sont indiqués dans la troisième colonne.

Le génotypage de ces marqueurs microsatellites a reposé sur la technologie des séquenceurs automatiques utilisant des amorces fluorescentes. Brièvement, après amplification, les produits de réaction de polymérisation en chaîne (PCR) fluorescents ont été déposés sur un gel de polyacrylamide sur séquenceur automatique selon les recommandations du constructeur (Perkin Elmer). La taille des allèles pour chaque sujet a été déduite grâce au logiciels Genescan<sup>R</sup> et Genotyper<sup>R</sup>. Les données ont ensuite été conservées sur une base informatique intégrée contenant les données généalogiques, phénotypiques et génétiques. Elles ont alors été utilisées pour les analyses de liaison génétique.

Plusieurs contrôles qualité ont été réalisés tout au long de la procédure de génotypage:

- 20      - double lecture indépendante des données de génotypage,
- utilisation d'un ADN standard servant de contrôle interne pour chaque migration électrophorétique,
- contrôle de la gamme de taille de chaque allèle observé,
- recherche d'erreurs de transmission mendéienne ,
- 25      - calcul de la distance génétique entre marqueurs (programme CRIMAP) et comparaison de celle-ci avec les données de la littérature,
- nouveau typage des marqueurs pour lesquels il était observé une recombinaison entre marqueurs proches.

Les données de génotypage ont été analysées par des méthodes de liaison génétique multipoint non paramétrique (Programme GENEHUNTER version 1.3). L'informativité du système de marqueurs était supérieure à 80% pour la région étudiée. Le maximum du test (NPL= 3,33; P = 0,0004) a été obtenu pour les marqueurs D16S541, D16S3117, D16S770 et D16S416 (figure 2).

Les données de typage pour ces 26 marqueurs de polymorphisme ont aussi été analysées à la recherche d'un déséquilibre de transmission. Deux groupes de 108 et 76 familles avec un ou plusieurs malades atteints de MC ont été étudiés. Le test statistique de déséquilibre de transmission a été décrit par Spielman et al. (1993). Il 5 n'a été pris en compte dans ce travail qu'un seul malade par famille et la valeur de p a été corrigée par le nombre d'allèles testés pour chaque marqueur étudié.

Un déséquilibre de transmission a été observé pour les allèles 4 et 5 (taille 205, resp. 207 paires de bases) du marqueur D16S3136 ( $p=0,05$ , resp.  $p=0,01$ ).

Ces résultats suggestifs d'une association entre le marqueur D16S3136 et la 10 MC ont conduit à construire une cartographie physique de la région génétique centrée sur D16S3136 et à établir la séquence d'un segment d'ADN génomique de grande taille (BAC) contenant ce site polymorphe. Il a alors été possible d'identifier et d'analyser un plus grand nombre de marqueurs de polymorphisme dans le voisinage de D16S3136 ainsi que de définir et d'étudier les séquences transcrives 15 présentes dans la région.

#### Exemple 2 : cartographie physique de la région IBD1

Un contig de fragments d'ADN génomique, centré sur les marqueurs D16S3136, D16S3117, D16S770 et D16S416, a été généré à partir des banques 20 d'ADN génomique humain de la fondation Jean Dausset/CEPH. Les segments d'ADN chromosomique ont été identifiés à partir de certains marqueurs de polymorphisme utilisés dans la cartographie génétique fine (D16S411, D16S416, D16S541, D16S770, D16S2623, D16S3035, D16S3117 et D16S3136). Pour chaque marqueur, une banque de chromosomes artificiels de bactéries (BAC) a été criblée 25 par PCR à la recherche de clones contenant la séquence du marqueur. Selon que les séquences testées étaient ou non présentes sur les clones de BAC il a été alors possible d'organiser les clones entre eux à l'aide du logiciel Segmap version 3.35.

On a pu établir, pour les BACs, une organisation continue (contig) couvrant la région génétique d'intérêt, selon une méthode connue de l'homme du métier 30 (Rouquier et al., 1994 ; Kim et al., 1996 ; Asakawa et al., 1997). Pour ce faire, les extrémités des BACs identifiés ont été séquencées et ces nouvelles données de séquence ont alors servi à cribler itérativement les banques de BACs. A chaque criblage, le contig de BAC a alors progressé d'un pas jusqu'à l'obtention d'un

continuum de clones chevauchants. La taille de chaque BAC participant au contig a été déduite de son profil de migration sur gel d'agarose en champ pulsé.

On a ainsi construit un contig de BAC contenant 101 BACs et s'étendant sur une distance globale de plus de 2,5 Mb avec une redondance moyenne de 5,5 BAC 5 à chaque point du contig. La taille moyenne des BAC est de 136kb.

Exemple 3 : séquençage du BAC hb87b10

Le BAC de ce contig contenant le marqueur de polymorphisme D16S3136 (appelé hb87b10), dont la taille était de 163761 bp a été séquencé selon la méthode 10 dite du "coup de fusil". En bref, l'ADN du BAC a été fragmenté par sonication. Les fragments d'ADN ainsi générés ont été soumis à une électrophorèse en gel d'agarose et ceux dont la taille était supérieure à 1,5 kb ont été élus pour être analysés. Ces fragments ont ensuite été clonés dans le phage m13 lui même introduit dans des bactéries rendues compétentes par électroporation. Après culture, l'ADN des clones 15 a été récupéré et séquencé par des méthodes de séquençage automatique à l'aide d'amorces fluorescentes du vecteur m13 sur séquenceur automatique.

1526 séquences différentes d'une taille moyenne de 600 bp ont été générées, qui ont été organisées entre elles grâce au logiciel Polyphredphrap<sup>R</sup> aboutissant à un contig de séquence couvrant l'ensemble du BAC. La séquence ainsi générée avait 20 une redondance moyenne de 5,5 équivalents génomiques. Les rares (n=5) intervalles de séquence non représentés dans la banque de clones m13 ont été comblés en générant des amorces de PCR spécifiques, de part et d'autre de ces intervalles, et en analysant le produit de PCR dérivé de l'ADN génomique d'un sujet sain.

25 Des homologies de séquence avec des séquences disponibles dans les bases de données génétiques publiques (Genbank) ont été recherchées. Aucun gène connu n'a pu être identifié dans cet intervalle de 163 kb. Plusieurs EST ont été positionnés suggérant que des gènes inconnus étaient contenus dans cette séquence. Ces EST issus des bases de données génétiques publiques (Genbank, GDB, Unigene, dbEST) 30 portaient les références suivantes : AI167910, AI011720, Rn24957, Mm30219, hs132289, AA236306, hs87296, AA055131, hs151708, AA417809, AA417810, hs61309, hs116424, HUMGS01037, AA835524, hs105242, SHGC17274, hs146128, hs122983, hs87280 et hs135201. La recherche d'exons putatifs à l'aide

du programme informatique GRAIL a permis d'identifier plusieurs exons potentiels, sites de polyadénylation et séquences promotrices.

Exemple 4 : études de déséquilibre de transmission

5 12 marqueurs de polymorphisme bialléliques (SNP) ont été identifiés dans une région s'étendant sur environ 250 kb et centrée sur le BAC hb87b10. Ces polymorphismes ont été générés par analyse de la séquence d'une dizaine de malades indépendants atteints de MC. Le séquençage a été le plus souvent réalisé au niveau d'EST connus et positionnés sur le BAC ou à son voisinage. Des exons 10 putatifs, prédits par le programme informatique GRAIL ont aussi été analysés. Les caractéristiques des marqueurs polymorphes ainsi identifiés sont rapportées sur le tableau 2.

Tableau 2. Caractéristiques de marqueurs de polymorphisme bialléliques étudiés

15 dans la région de IBD1

| I | II              | III      | IV            | V                 | VI  |
|---|-----------------|----------|---------------|-------------------|-----|
| 1 | KIAA0849ex9     | PCR-AS   |               | SEQ ID N° 88 à 90 | 116 |
| 2 | hb27G11F        | PCR-RFLP | <i>Bsr</i> I  | SEQ ID N° 86, 87  | 185 |
|   |                 |          |               |                   | 116 |
|   |                 |          |               |                   | 69  |
| 3 | Ctg22Ex1        | PCR-RFLP | <i>Rsa</i> I  | SEQ ID N° 84, 85  | 381 |
|   |                 |          |               |                   | 313 |
|   |                 |          |               |                   | 69  |
| 4 | SNP1            | PCR-AS   |               | SEQ ID N° 81 à 83 | 410 |
| 5 | ctg2931-3ac/ola | LO       |               | SEQ ID N° 78 à 80 | 51  |
|   |                 |          |               |                   | 49  |
| 6 | ctg2931-5ag/ola | LO       |               | SEQ ID N° 75 à 77 | 44  |
|   |                 |          |               |                   | 42  |
| 7 | SNP3-2931       | PCR-AS   |               | SEQ ID N° 72 à 74 | 245 |
| 8 | Ctg25Ex1        | PCR-RFLP | <i>Bst</i> II | SEQ ID N° 70, 71  | 207 |
|   |                 |          |               |                   | 122 |
|   |                 |          |               |                   | 85  |

|    |            |          |             |                    |                  |
|----|------------|----------|-------------|--------------------|------------------|
| 9  | CTG35 ExA  | PCR-AS   |             | SEQ ID N° 67 à 69  | 333              |
| 10 | ctg35 ExC  | PCR-AS   |             | SEQ ID N° 64 à 66  | 198              |
| 11 | D16S3136   |          |             | SEQ ID N° 37, 38   |                  |
| 12 | hb133D1f   | PCR-RFLP | <i>TaqI</i> | - SEQ ID N° 62, 63 | 369<br>295<br>74 |
| 13 | D16S3035   |          |             | SEQ ID N° 35, 36   |                  |
| 14 | ADCY7 int7 | PCR-AS   |             | SEQ ID N° 59 à 61  | 140              |

PCR-AS : PCR-allèle spécifique ; LO : Ligature d'oligonucléotides

Les 12 marqueurs de polymorphisme bialléliques nouvellement décrits dans ce travail sont répertoriés dans ce tableau. Pour chacun d'eux sont indiqués :

5            - le locus (colonne I)  
               - le nom (colonne II)  
               - la technique de génotypage utilisée (colonne III)  
               - l'enzyme de restriction éventuellement utilisée (colonne IV)  
               - les amores oligonucléotidiques utilisées pour la réaction de  
 10            polymérisation en chaîne ou pour la ligature (colonne V)  
               - la taille des produits attendus lors du typage (colonne VI)

199 familles comportant 1 ou plusieurs malades atteints de MC ont été typées pour ces 12 marqueurs de polymorphisme ainsi que pour les marqueurs D16S3035 et D16S3136 localisés sur le BAC hb87b10. Les familles comportant des  
 15            malades atteints de RCH n'ont pas été prises en compte. Les méthodes de typage des polymorphismes étudiés ont été variables en fonction du type de polymorphisme faisant appel à :

20            - la technique de PCR-RFLP (amplification suivie de digestion enzymatique du produit de PCR) quand le polymorphisme était situé sur un site de restriction enzymatique.  
               - PCR avec amores spécifiques du site polymorphe : amplification différentielle des deux allèles en utilisant des amores spécifiques de chaque allèle.

- Test de ligation d'oligonucléotides : ligation différentielle utilisant des oligonucléotides spécifiques de chaque allèle, suivie d'électrophorèse en gel de polyacrylamide.

Les données de typage ont ensuite été analysées selon un test de déséquilibre de transmission (programme informatique TDT du logiciel GENEHUNTER version 2). Pour les familles comportant plusieurs apparentés atteints, un seul malade a été pris en compte pour l'analyse. En effet, la prise en compte de plusieurs malades apparentés pose le problème de non indépendance des données dans les calculs statistiques et peut induire une inflation de la valeur du test. Le malade servant à l'analyse a été tiré au sort au sein de chaque famille par une procédure automatique de randomisation. Compte tenu de cette randomisation, la valeur du test statistique obtenu ne représentait qu'un seul échantillon possible issu du groupe de familles étudiées. Afin de ne pas limiter l'analyse à ce seul échantillon possible et pour mieux appréhender la robustesse des résultats obtenus, pour chaque test, une centaine d'échantillons aléatoires ont ainsi été générés et analysés.

Les marqueurs ont été étudiés séparément puis groupés selon leur ordre sur le segment chromosomal (KIAA0849ex9 (locus 1), hb27G11F (locus 2), Ctg22Ex1 (locus 3), SNP1 (locus 4), ctg2931-3ac/ola (locus 5), ctg2931-5ag/ola (locus 6), SNP3-2931 (locus 7), Ctg25Ex1 (locus 8), CTG35ExA (locus 9), ctg35ExC (locus 10), d16s3136 (locus 11), hb133D1f (locus 12), D16S3035 (locus 13), ADCY7int7 (locus 14)) (tableau 2). Les haplotypes comportant 2, 3 et 4 marqueurs consécutifs ont ainsi été analysés en utilisant toujours la même stratégie (100 échantillons aléatoires en prenant pour chaque famille un seul individu atteint).

Pour chaque échantillon testé, il n'a été pris en compte que les génotypes (ou haplotypes) portés par au moins 10 chromosomes parentaux. En moyenne 250 tests différents ont ainsi été réalisés pour chaque échantillon. Il a alors été possible de déduire le nombre de tests attendus positifs pour chaque seuil de signification et de comparer cette distribution à la distribution observée. Pour les sujets sains, la distribution des tests n'est pas différente de celle attendue selon le hasard ( $\chi^2 = 2,85$ , ddl=4, p=0,58). Pour les sujets malades, au contraire, il existe un excès de tests positifs témoignant de l'existence d'un déséquilibre de transmission dans la région étudiée.

Les résultats des tests de déséquilibre de transmission pour chaque marqueur de polymorphisme pris isolément et pour les haplotypes montrant les plus forts déséquilibres de transmission ont montré que les marqueurs suivants sont en déséquilibre de liaison avec la maladie: Ctg22Ex1 (locus 3), SNP1 (locus 4), 5 ctg2931-5ag/ola (locus 6), SNP3-2931 (locus 7), Ctg25Ex1 (locus 8) et ctg35ExC (locus 10). Ces marqueurs s'étendent sur une région d'environ 50kb (positions 74736 à 124285 sur la séquence de hb87b10).

Les haplotypes les plus fortement associés avec la maladie de Crohn s'étendent eux aussi sur cette région. Ainsi, pour la majorité des échantillons 10 aléatoires, le test de transmission était positif ( $p < 0,01$ ) pour des haplotypes combinant les marqueurs suivants :

- locus 5-6, locus 6-7, locus 7-8, locus 8-9, locus 9-10, locus 10-11
- locus 5-6-7, locus 6-7-8, locus 7-8-9, locus 8-9-10, locus 9-10-11
- locus 5-6-7-8; locus 6-7-8-9, locus 7-8-9-10,

15 L'haplotype de susceptibilité le plus à risque est défini par les locus 7 à 10. Il s'agit de l'haplotype 1-2-1-2 (tableau 2).

Les marqueurs testés sont, comme attendu, le plus souvent en déséquilibre de liaison entre eux.

Plus récemment, un nouveau test, le Pedigree Disequilibrium Test (PDT), 20 publié en juillet 2000 (Martin et al., 2000) a été utilisé pour mieux appréhender la signification des résultats obtenus avec le programme informatique TDT. Cette nouvelle statistique permet en effet d'utiliser l'ensemble de l'information disponible dans une famille, tant à partir des sujets malades qu'à partir des sujets sains et de pondérer l'importance de chaque apparenté en une statistique globale pour chaque 25 famille. Les valeurs de  $p$  correspondant aux tests PDT et obtenues pour un groupe élargi de 235 familles avec un ou plusieurs apparentés atteints de la maladie de Crohn sont rapportées dans le Tableau 3. Cette nouvelle analyse confirme que la région du BAC hb87b10 est bien associée avec la maladie de Crohn.

Tableau 3. Résultat des tests PDT réalisés sur 235 familles atteintes de la maladie de Crohn (NS : non significatif)

| LOCUS           | VALEUR p DU TEST PDT |
|-----------------|----------------------|
| KIAA0849ex9     | NS                   |
| hb27g11f        | 0,05                 |
| ctg22ex1        | 0,01                 |
| SNP1            | 0,001                |
| ctg2931-3ac/ola | NS                   |
| ctg2931-5ag/ola | 0,0001               |
| SNP3-2931       | 0,0001               |
| ctg25ex1        | 0,0006               |
| ctg35exA        | NS                   |
| ctg35exC        | 0,00002              |
| D16S3136        | NS                   |
| hb133d1f        | NS                   |
| D16S3035        | NS                   |

Exemple 5 : Identification du gène IBD1

5 Les groupements d'EST (références Unigene : Hs 135201, Hs87280, Hs122983, Hs146128, Hs105242, Hs116424, Hs61309, Hs151708, Hs 87296 et Hs132289) publiés et présents sur le BAC hb87b10 ont été étudiés à la recherche d'une séquence d'ADN complémentaire (ADNc) plus complète. Pour IBD1prox, les clones disponibles dans les banques publiques ont été séquencés et les séquences organisées entre elles. Pour IBD1, une banque d'ADN complémentaire de sang périphérique (Stratagene human blood cDNA lambda zapexpress ref 938202) a été criblée par les produits de PCR générés à partir des EST connus selon les modalités proposées par le fabriquant. La séquence des ADNc ainsi identifiés a ensuite servi à un nouveau criblage de la banque d'ADNc et ainsi de suite jusqu'à l'obtention de

10

15 l'ADNc présenté.

L'EST hs135201 (UniGene) a permis d'identifier un ADNc ne figurant pas sur les bases de données génétiques disponibles (Genbank). Il correspond donc à un nouveau gène humain. La comparaison de la séquence du cDNA et de l'ADN génomique a montré que ce gène est constitué de 11 exons et 10 introns. Un exon

supplémentaire, en position 5' par rapport au cDNA identifié est prédit par l'analyse de la séquence avec le logiciel Grail. Ces exons sont très homologues avec les premiers exons du gène CARD4/NOD1. Considérant l'ensemble des exons identifiées et l'exon putatif supplémentaire, ce nouveau gène apparaît avoir une 5 structure génomique très proche de celle de CARD4/NOD1. Par ailleurs, en amont du premier exon putatif figure un site d'initiation de la transcription. Pour l'ensemble de ces raisons, l'exon putatif a été considéré comme participant à ce nouveau gène. L'ADNc reporté en annexe (SEQ ID N° 1) comporte donc l'ensemble de la séquence identifiée plus la séquence prédicta par la modélisation informatique, 10 l'ADN complémentaire débutant arbitrairement au premier codon ATG de la séquence codante prédicta. Sur cette base, le gène comporterait donc 12 exons et 11 introns. La structure intron-exon du gène est rapportée sur la SEQ ID N° 3.

La séquence protéique déduite de la séquence nucléotidique, comporte 1041 acides aminés (SEQ ID N° 2). Cette séquence n'a pas non plus été retrouvée sur les 15 bases de données biologiques (Genpept, pir, swissprot).

Or, plus récemment, l'exon putatif ci-dessus décrit n'a pas pu être confirmé. Le gène IBD1 ne comporte donc effectivement que 11 exons et 10 introns et code pour une protéine de 1013 acides aminés (c'est-à-dire 28 acides aminés de moins que déterminé initialement).

20 L'étude de la séquence protéique déduite montre que ce gène contient trois domaines fonctionnels différents (figure 3) :

- Un domaine CARD (Caspase Recruitment Domain) connu pour être impliqué dans l'interaction entre protéines régulatrices de l'apoptose et de l'activation de la voie NFkappa B. Le domaine CARD permet de classer cette nouvelle protéine dans la famille des protéines CARD dont les membres les plus anciens sont CED 4, APAF1 et RICK.
- Un domaine NBD (Nucléotide Binding Domaine) comportant un site de reconnaissance de l'ATP et un site de liaison du Magnésium. La protéine doit donc avoir une activité kinase très probable.
- Un domaine LRR (Leucine Rich Domain) supposé participer à l'interaction entre protéines par analogie avec d'autres domaines protéiques décrits .

Par ailleurs, le domaine LRR de la protéine permet d'affilier la protéine à une famille de protéines impliquées dans la signalisation intracellulaire et présentes tant chez les plantes que chez les animaux.

La comparaison de ce nouveau gène avec les gènes précédemment identifiés 5 et disponibles dans les bases de données publiques montre que celui-ci est très homologue avec CARD4/NOD1 (Bertin et al., 1999 ; Inohara et al., 1999). Cette homologie porte sur la séquence de l'ADN complémentaire, la structure intron-exon du gène et la séquence protéique. L'identité de séquence des 2 ADN complémentaires est de 58%. Une similitude est également observée au niveau de la 10 structure introns-exons. L'homologie de séquence au niveau protéique est de l'ordre de 40%.

La similitude entre ce nouveau gène et CARD4/NOD1 suggère que, comme CARD4/NOD1, la protéine IBD1 est impliquée dans la régulation de l'apoptose et de l'activation de NF-kappa B (Bertin et al., 1999 ; Inohara et al., 1999). La 15 régulation de l'apoptose cellulaire et l'activation de NF-kappa B sont des voies de signalisation intracellulaire essentielles dans les réactions immunitaires. En effet, ces voies de transduction du signal sont les voies effectrices des protéines de la famille du récepteur du TNF (Tumor Necrosis Factor) impliquées dans les interactions cellule-cellule et la réponse cellulaire aux différents médiateurs de 20 l'inflammation (cytokines). Le nouveau gène apparaît donc comme potentiellement important à la réaction inflammatoire, de façon générale.

Plusieurs faisceaux de preuves viennent à l'appui de la dérégulation de NF- 25 kB induit par des bactéries dans la maladie de Crohn. Tout d'abord, la susceptibilité à IBD spontanée chez les souris a été associée à des mutations dans Tlr4, une molécule connue pour se lier aux LPS par l'intermédiaire de son domaine LRR (Poltorak et al., 1998 et Sundberg et al., 1994) et pour être un membre des activateurs de la famille de NF-kB. Deuxièmement, la thérapie antibiotique cause une amélioration provisoire chez les patients atteints de MC accréditant l'hypothèse que les bactéries entériques peuvent jouer un rôle étiologique dans la maladie de 30 Crohn (McKay, 1999). Troisièmement, NF-kB joue un rôle pivot dans les maladies inflammatoires de l'intestin et est activé dans les cellules mononucléées de la lamina propria dans la maladie de Crohn (Schreiber et al., 1998). Quatrièmement, le traitement de la maladie de Crohn est basée sur l'utilisation de la sulfasalazine et

des glucocorticoïdes, tous deux connus comme étant des inhibiteurs de NF- $\kappa$ B (Auphan et al., 1995 et Wahl et al., 1998)

Encore plus récemment, il a été montré que le gène candidat IBD1 code pour une protéine très similaire à NOD2, un membre de la superfamille 5 CED4/APAF1 (Ogura et al., 2000). Les séquences nucléotidiques et protéiques de IBD1 et NOD2 ne divergent en réalité que pour une petite portion toute initiale des 2 séquences rapportées. Les expressions tissulaires de Nod2 et IBD1 sont de plus superposables. Ces deux gènes (protéines) peuvent donc être considéré(e)s comme identiques. Il a été démontré que le domaine LRR de Nod2 a une activité de liaison 10 pour les lipopolysaccharides bactériens (LPS) (Inohara et al., 2000) et que sa délétion stimule la voie de NF $\kappa$ B. Ce résultat confirme les données de l'invention.

L'expression tissulaire de IBD1 a été ensuite étudiée par la technique du Northern Blot. Un transcrit de 4.5 kb est visible dans la plupart des tissus humains. La taille du transcrit est conforme avec la taille prédictive par l'ADNc. Le transcrit de 15 4.5 kb semble en très faible abondance dans l'intestin grêle et le colon. Il est par contre très fortement exprimé dans les globules blancs. Ceci est en accord avec des données cliniques sur les transplantations qui suggèrent que la maladie de Crohn est potentiellement une maladie liée aux cellules immunitaires circulantes. En effet, la transplantation intestinale n'empêche pas la récidive sur le greffon dans la maladie 20 de Crohn tandis que la transplantation de moelle osseuse semble avoir un effet bénéfique sur l'évolution de la maladie.

Certaines données font également penser à un épissage alternatif, qui pourrait s'avérer un élément important dans la possibilité de générer des mutants qui pourraient jouer un rôle dans le développement de maladies inflammatoires.

25 Le promoteur du gène IBD1 n'est actuellement pas identifié avec précision. Il est cependant raisonnable de penser, par analogie avec un très grand nombre de gènes que celui-ci réside, au moins pour partie, immédiatement en amont du gène, dans la portion 5' de celui-ci. Cette région génétique contient des séquences transcrtes comme en témoigne la présence d'EST (HUMGS01037, AA835524, 30 hs.105242, SHGC17274, hs.146128, hs.122983, hs.87280). Les clones ATCC contenant ces séquences ont été séquencés et analysés dans le laboratoire, permettant de mettre en évidence une organisation en exons et en introns avec d'éventuels épissages alternatifs. Ces données suggèrent l'existence d'un autre gène

(nommé IBD1prox en raison de sa proximité d'IBD1). La séquence partielle de l'ADN complémentaire de IBD1prox est rapportée (SEQ ID N° 4) de même que sa structure intron-exon sur la SEQ ID N° 6.

La traduction des ADNc correspondant à IBD1prox aboutit à une protéine 5 contenant une homéobox. L'analyse de plusieurs ADNc du gène suggère cependant l'existence d'épissages alternatifs. IBD1prox, selon un des épissages alternatifs possibles correspond à l'EST anonyme HUMGS01037 dont l'ARN est exprimé de manière plus importante dans les lignées leucocytaires différenciées que dans les lignées non différenciées.

10 Ainsi, il est possible que ce gène puisse avoir un rôle dans l'inflammation et la différentiation cellulaire. Il peut donc lui aussi être considéré comme un bon candidat pour la susceptibilité aux MICI. L'association entre MC et le polymorphisme ctg35 ExC localisé sur la séquence codante de IBD1prox renforce cette hypothèse même si ce polymorphisme n'entraîne pas de variation de séquence 15 au niveau protéique.

Enfin, plus récemment, l'existence d'une liaison génétique dans les familles atteintes de la maladie de Crohn et ne comportant pas de mutation du gène IBD1 suggère elle aussi que IBD1 prox a un rôle additionnel à IBD1 dans la prédisposition génétique à la maladie.

20 La relation fonctionnelle entre IBD1 et IBD1prox n'est actuellement pas établie. Toutefois, la forte proximité entre les deux gènes pourrait refléter une interaction entre ceux-ci. Dans ce cas, la localisation « tête -bêche » de ces gènes suggère qu'ils puissent avoir des modes de régulation communs ou interdépendants.

25 Exemple 6 : identifications de mutations du gène IBD1 dans les maladies inflammatoires

Afin de confirmer le rôle de IBD1 dans les maladies inflammatoires, la séquence codante et les jonctions intron-exon du gène ont été séquencées de l'exon 2 à l'exon 12 inclus chez 70 sujets indépendants, à savoir : 50 malades atteints de 30 MC, 10 malades atteints de RCH, 1 malade atteint de syndrome de Blau et 9 témoins sains. Les malades étudiés étaient pour la plupart des formes familiales de la maladie et étaient souvent porteurs de l'haplotype de susceptibilité défini par les

études de déséquilibre de transmission. Les témoins sains étaient d'origine caucasienne.

24 variants de séquence ont ainsi pu être identifiés sur ce groupe de 70 personnes non apparentées(tableau 3).

5 La nomenclature des mutations rapportées fait référence à la séquence initiale de la protéine comportant 1041 acides aminés. La nomenclature plus récemment proposée est aisément déduite en retirant 28 acides aminés à la séquence initiale, et correspond donc à une protéine comprenant 1013 acides aminés (cf exemple 5).

10

Tableau 4. Mutations observées dans le gène IBD1

| Exon | Variant nucléotidique | Variant protéique | Maladie de Crohn | Rectocolite hémorragique | Témoins sains |
|------|-----------------------|-------------------|------------------|--------------------------|---------------|
| 1    | non testé             |                   |                  |                          |               |
| 2    | G417A                 | silencieux        |                  |                          |               |
| 2    | C537G                 | silencieux        |                  |                          |               |
| 3    | aucun                 |                   |                  |                          |               |
| 4    | T805C                 | S269P             | 48/100           | 6/20                     | 3/18          |
| 4    | A869G                 | N290S             | 0                | 0                        | 1/18          |
| 4    | C905T                 | A302V             | 1/100            | 0                        | 0             |
| 4    | C1283T                | P428L             | 1/100            | 0                        | 0             |
| 4    | C1284A                | silencieux        |                  |                          |               |
| 4    | C1287T                | silencieux        |                  |                          |               |
| 4    | T1380C                | silencieux        |                  |                          |               |
| 4    | T1764G                | silencieux        |                  |                          |               |
| 4    | G1837A                | A613T             | 1/100            | 0                        | 0             |
| 4    | C2107T                | R703W             | 10/10            | 1/20                     | 1/18          |
| 4    | C2110T                | R704C             | 4/10             | 1/20                     | 0             |
| 5    | G2365A                | R792Q             | 1/100            | 0                        | 0             |
| 5    | G2370A                | V794M             | 0                | 1/20                     | 0             |
| 5    | G2530A                | E844K             | 1/10             | 0                        | 0             |
| 6    | A2558G                | N853S             | 1/100            | 0                        | 0             |
| 6    | A2590G                | M864V             | 1/100            | 0                        | 0             |
| 7    | aucun                 |                   |                  |                          |               |
| 8    | G2725C                | G909R             | 7/100            | 0                        | 0             |
| 8    | C2756A                | A919D             | 1/100            | 0                        | 0             |
| 9    | G2866A                | V956I             | 2/100            | 1/20                     | 3/18          |
| 10   | C2928T                | silencieux        |                  |                          |               |
| 11   | 3022insC              | stop              | 20/100           | 0                        | 0             |

|    |       |  |  |  |  |
|----|-------|--|--|--|--|
| 12 | aucun |  |  |  |  |
|----|-------|--|--|--|--|

Les mutations autres que silencieuses observées dans chaque exon sont rapportées. Elles sont indiquées par la variation de la chaîne peptidique. Pour chaque mutation et pour chaque phénotype étudié, il est indiqué le nombre de fois où la mutation est observé, rapporté au nombre de chromosomes testés.

5 Aucun variant de séquence fonctionnel n'a été identifié dans les exons 1 à 3 (correspondants au domaine CARD de la protéine). Les exons 7 et 12 n'ont pas non plus montré de variation de séquence. Certains variants correspondaient à des polymorphismes déjà identifiés et typés pour les études de déséquilibre de transmission, à savoir :

10 -Snp3-2931 : variant nucléotidique T805C, variant protéique S269P  
 -ctg2931-5ag/ola : variant nucléotidique T1380C (silencieux)  
 -ctg2931-3ac/ola : variant nucléotidique T1764G (silencieux)  
 -SNP1 : variant nucléotidique C2107T, variant protéique R703W

Plusieurs variations de séquence étaient silencieuses (G417A, C537G, 15 C1284A, C1287T, T1380C, T1764G, C2928T) et n'entraînaient pas de modification de la séquence protéique. Elles n'ont pas été étudiées davantage ici.

Pour les 16 variations de séquence non silencieuses, il a été observé des variants de séquence protéique chez 43/50 MC contre 5/9 témoins sains et 6/10 RCH. L'existence d'une ou plusieurs variation(s) de séquence apparaissait associée 20 au phénotype MC. Il existait souvent plusieurs variations de séquence chez un même individu atteint de MC suggérant un effet parfois récessif du gène pour la MC. A l'inverse, aucun homozygote ou hétérozygote composite n'était observé parmi les patients atteints de RCH ou parmi les témoins sains.

Certains variants non silencieux étaient présents à la fois chez les malades 25 atteints de RCH ou de MC et chez les sujets sains. Il s'agissait des variants S269P, N290S, R703W et V956I situés dans les exons 2, 4 et 9. Un complément d'information semble donc nécessaire avant de retenir un éventuel rôle fonctionnel à ces variants de séquence.

V956I est une variation de séquence conservative (acides aminés 30 aliphatiques).

Le variant de séquence S269P correspond à une variation de classe d'acide aminé (hydroxylé en immunoacide) au début du domaine liant les nucléotides. Il en

déséquilibre de transmission avec la MC. Il s'agit en effet du polymorphisme Snp3 (Cf. supra).

R703W aboutit à une modification de la classe de l'acide aminé (aromatique au lieu de basique). Cette modification survient dans la région intermédiaire entre 5 les domaines NBD et LRR, région conservée entre IBD1 et CARD4/NOD1. Un rôle fonctionnel peut donc être suspecté pour ce polymorphisme. Cette variation de séquence (correspondant au site polymorphe Snp1) est plus souvent transmise au malades atteints de MC que ne le veut le hasard (Cf. supra) confirmant que ce polymorphisme est associé à la MC. Il est possible que la présence de ce mutant 10 chez les sujets sains témoigne d'une pénétrance incomplète de la mutation comme cela est attendu pour les maladies génétiques complexes telles que les maladies inflammatoires chroniques de l'intestin.

Le variant R704C, situé immédiatement à coté de R703W a pu être identifié à la fois dans la MC et dans la RCH. Il correspond lui aussi à une variation non 15 conservative de la protéine (acide aminé soufré au lieu de basique) sur la même région protéique, suggérant un effet fonctionnel aussi important pour R704C que pour R703W.

D'autres variations de séquence sont spécifiques de la MC de la RCH ou du syndrome de Blau.

20 Certaines variations de séquence sont au contraire rares, présentes chez un ou quelques malades (A613T, R704C, E844K, N853S, M864V, A919D). Il s'agit toujours de variations entraînant des modifications non conservatives de la protéine dans des domaines leucine riches, à des positions importantes au sein de ces domaines. Ces différents éléments suggèrent que ces variations ont un rôle 25 fonctionnel.

Deux variations de séquence (G909R, L1008P\*) sont retrouvées chez un assez grand nombre de malades de Crohn (respectivement 7/50 et 16/50) alors qu'elles ne sont pas détectées chez les témoins ou chez les malades atteints de RCH.

La délétion/insertion d'une guanosine au niveau du codon 1008 aboutit à une 30 transformation de la troisième leucine de l'hélice alpha du dernier LRR en proline suivie d'un codon STOP (L1008P\*). Cette variation de séquence entraîne donc une modification importante de la protéine : réduction de taille de la protéine (protéine possédant un domaine LRR tronqué) et altération d'un acide aminé très conservé

(Leucine). Cette modification de séquence est associée à la MC comme en témoigne une étude de déséquilibre de transmission dans 16 familles porteuses de la mutation (P=0,008).

La mutation G909R survient sur le dernier acide aminé du sixième motif 5 LRR. Il remplace un acide aminé aliphatique en acide aminé basique. Cette variation est potentiellement importante compte tenu du caractère habituellement neutre ou polaire des acides aminés en position terminale des motifs leucine riche (tant pour IBD1 que pour NOD1/CARD4) et du caractère conservé de cet acide aminé sur les protéines IBD1 et NOD1/CARD4.

10 Dans le syndrome de Blau, les malades (n=2) de la famille étudiée étaient porteurs d'une variation de séquence spécifique (L470F), localisée dans l'exon 4 et correspondant au domaine NBD de la protéine. Dans cette série, ce variant de séquence était spécifique du syndrome de Blau.

15 Dans la RCH, plusieurs variants de séquence non retrouvés chez les sujets sains ont aussi été identifiés. La proportion de malades porteurs d'une mutation était plus modeste que pour la MC, comme attendu compte tenu de la liaison moins fortement établie entre IBD1 et RCH et du caractère supposé moins génétique de cette dernière maladie. Des variations de séquence étaient communes à la MC et à la RCH (R703W, R704C). D'autres au contraires apparaissaient spécifiques de la 20 RCH (V794M). Cette observation permet de confirmer que MC et RCH sont des maladies partageant au moins en partie la même prédisposition génétique. Elle pose les bases d'une classification nosologique des MICI.

25 L'étude des variants de séquence du gène IBD1 a donc permis d'identifier plusieurs variants ayant un effet fonctionnel très probable (ex : protéine tronquée) et associés à la maladie de Crohn, à la RCH et au syndrome de Blau.

Le promoteur du gène n'est actuellement pas déterminé. Selon toute vraisemblance cependant, celui-ci est probablement situé dans la région 5' en amont du gène. Selon cette hypothèse, les variants de séquence observés dans cette région peuvent avoir un effet fonctionnel. Ceci pourrait expliquer la très forte association 30 entre MC et certains locus polymorphes tels que ctg35 ExC ou Ctg25Ex1.

L'invention fournit ainsi la première description de mutations dans la famille des gènes contenant un domaine CARD chez l'homme. La fréquence de ces mutations dans des maladies inflammatoires variées montre que le gène IBD1 a un

rôle essentiel dans le processus inflammatoire normal et pathologique. Cette invention fournit de nouvelles voies de compréhension et de recherche dans le domaine de la physiopathologie des processus inflammatoires normaux et pathologiques. Elle permet de ce fait d'envisager le développement de nouvelles 5 molécules pharmaceutiques régulant les voies effectrices contrôlées par IBD1 et utiles dans le traitement des maladies inflammatoires et la régulation du processus inflammatoire en général.

Exemple 7 : bases d'un diagnostic biologique de susceptibilité à la maladie de 10 Crohn

Plus récemment, 457 patients indépendants atteints de la maladie de Crohn, 159 patients indépendants atteints de rectocolite hémorragique et 103 témoins sains ont été étudiés à la recherche de mutations. Ce travail a permis de confirmer les mutations précédemment rapportées et d'identifier des mutations supplémentaires 15 rapportées sur la figure 4. Les mutations principales ont ensuite été génotypées dans 235 familles atteintes de la maladie de Crohn. Ce travail plus récent est exposé en utilisant comme référence la séquence protéique plus courte (1013 acides aminés, voir exemple 5) mais la nomenclature antérieure des mutations est aisément déduite 20 à partir de cette dernière en ajoutant 28 au chiffre indiquant la position des acides aminés.

Parmi les 5 mutations les plus fréquentes, la mutation conservative V928I (anciennement V956I) n'est pas significativement associée à l'une ou l'autre des maladies inflammatoires de l'intestin et ne semble donc pas avoir de rôle important dans la maladie.

25 La mutation S241P (anciennement S269P) est en déséquilibre de liaison avec les autres mutations principales et ne semble pas jouer par elle-même un rôle important dans la susceptibilité aux maladies inflammatoires de l'intestin (données non montrées).

A l'inverse, les 3 autres mutations R675W (anciennement R703W), G881R 30 (anciennement G909R) et 980fs (anciennement L1008P\*) sont significativement associées à la maladie de Crohn mais pas à la rectocolite hémorragique (cf infra). La localisation dans le LRR ou à sa proximité immédiate des 3 mutations fréquentes plaide très fortement pour un mécanisme fonctionnel impliquant ce domaine

protéique, probablement par un défaut de régulation négative de NFkB par la protéine mutée. Les autres mutations sont plus rares (figure 4). Ces mutations cumulées sont présentes chez 17% des sujets atteints de la maladie de Crohn contre respectivement 4 % et 5 % les sujets sains ou atteints de rectocolite hémorragique.

5 Un grand nombre des mutations rares sont aussi localisées dans le LRR.

Les études intrafamiliales des trois polymorphismes les plus fréquents dans la maladie de Crohn montrent qu'ils sont tous trois associés à la maladie (tableau 5).

Comme attendu, pour une mutation supposée très délétère, le polymorphisme le plus fortement associé est la mutation tronquante. Ces trois polymorphismes sont

10 associés de manière indépendante à la maladie de Crohn puisqu'il n'a pas été possible d'identifier sur 235 familles des chromosomes porteurs de plus d'une de ces trois mutations. Le caractère indépendant de ces associations renforce considérablement l'hypothèse que le gène IBD1 est bien impliqué dans la prédisposition génétique à la maladie de Crohn.

15

Tableau 5 : étude des 3 polymorphismes fréquents de IBD1 dans 235 familles atteintes de la maladie de Crohn

| MUTATION | VALEUR p DU TEST PDT |
|----------|----------------------|
| R675W    | 0,001                |
| G881R    | 0,003                |
| 980fs    | 0,000006             |

20 Les études de cas-témoin confirment cette association (tableau 6). Ils montrent que les mutations les plus fréquentes dans la maladie de Crohn ne sont pas fréquentes dans la rectocolite hémorragique.

Tableau 6 : étude de cas-témoin des 3 polymorphismes fréquents de IBD1 dans les maladies inflammatoires de l'intestin

| MUTATION       | NB DE CHROMOSOMES STUDIES | FREQUENCE DE L'ALLELE A RISQUE R675W | FREQUENCE DE L'ALLELE A RISQUE G881R | FREQUENCE DE L'ALLELE A RISQUE 980fs | TOTAL ALLELES A RISQUE |
|----------------|---------------------------|--------------------------------------|--------------------------------------|--------------------------------------|------------------------|
| Témoins sains  | 206                       | 0,04                                 | 0,01                                 | 0,02                                 | 0,07                   |
| Rectocolite H. | 318                       | 0,03                                 | 0,00                                 | 0,01                                 | 0,05                   |

|         |     |      |      |      |      |
|---------|-----|------|------|------|------|
| M Crohn | 936 | 0,11 | 0,06 | 0,12 | 0,29 |
|---------|-----|------|------|------|------|

L'étude de l'effet dose de ces mutations montre que les sujets porteurs d'une mutation à l'état homozygote ou hétérozygote composite présentent un bien plus grand risque de développer la maladie que les sujets non porteurs ou hétérozygotes pour ces mutations (tableau 7).

Tableau 7 : risque relatif et absolu de la maladie de Crohn attribuable en fonction du génotype de IBD1

Dans la population générale, un risque de la maladie de Crohn de 0,001 a été pris comme référence et les mutations ont été supposées en équilibre de Hardy-Weinberg.

| DISTRIBUTION               | GENOTYPE      |                     |            |                        |
|----------------------------|---------------|---------------------|------------|------------------------|
|                            | AUCUN VARIANT | SIMPLE HETEROZYGOTE | HOMOZYGOTE | HETEROZYGOTE COMPOSITE |
| Sains                      | 88            | 15                  | 0          | 0                      |
| Rectocolite H              | 145           | 13                  | 1          | 0                      |
| M Crohn                    | 267           | 133                 | 28         | 40                     |
| Risque attribuable de MC : |               |                     |            |                        |
| Risque relatif             | 1             | 3                   | 38         | 44                     |
| Risque absolu              | 0,0007        | 0,002               | 0,03       | 0,03                   |

Les travaux cités ci-dessus confirment les données préliminaires antérieures et apportent les bases détaillées d'un diagnostic biologique de la maladie de Crohn par l'étude des variants de IBD1. En effet, ce travail :

- 1) définit les mutations dont la fréquence est supérieure à 0,001 dans une population caucasienne mélangée,
- 2) définit la fréquence des mutations observées et permet de définir 3 mutations principales associées à la maladie de Crohn. Ainsi, il est possible, grâce à ce travail, de définir une stratégie d'étude du gène pour la recherche de variants morbides à savoir : premièrement typage des 3 mutations principales, deuxièmement recherche de mutations dans les 7 derniers exons, troisièmement recherche d'autres variants de séquence.

3) définit les modalités pratiques de recherche de ces mutations en signalant leur position et leur nature. En effet, il est ensuite aisément à l'homme du métier de mettre au point des méthodes de typage et de séquençage selon son expertise personnelle. On peut citer en particulier la possibilité de faire les génotypages des 3 mutations principales par PCR suivie de digestion enzymatique et électrophorèse, étude des profils de migration par dHPLC, DGGE ou SSCP, oligoligation, microséquençage, etc.

5) démontre l'indépendance des mutations les plus fréquentes qui ne sont pas observées sur le même chromosome dans cette population étendue et variée. Cette information permet de classer de façon fiable les sujets en hétérozygotes composites (ayant deux mutations) comme porteur à une double dose de variations intragéniques.

10) démontre que la plus grande proportion des mutations n'entraîne qu'un effet nul ou minime sur le risque de rectocolite hémorragique. Ce résultat permet d'envisager d'aider le clinicien dans le diagnostic différentiel entre ces deux maladies. En effet, dans environ 10 % des cas, les maladies inflammatoires de l'intestin restent inclassées malgré les examens biologiques, radiologiques et endoscopiques.

15) définit un risque relatif et absolu de la maladie pour les génotypes les plus fréquents. Ce résultat pose les bases d'un diagnostic prédictif potentiellement utile dans une démarche de suivi ou d'intervention préventive dans les populations à risque, en particulier, les apparentés de malades.

20) démontre l'existence d'un effet dose pour le gène IBD1 et confirme le caractère en partie récessif de la prédisposition génétique à la maladie de Crohn. Il permet donc de poser les bases d'un conseil génétique et d'un diagnostic préclinique intrafamilial.

25) Notons enfin qu'une mutation supplémentaire du domaine NBD a été isolée dans une deuxième famille porteuse d'un syndrome de Blau. La rareté des deux événements dans 2 familles différentes suffit à confirmer l'implication de ce gène dans le syndrome de Blau et dans les maladies granulomateuses en générale.

L'ensemble de ces données apporte un outil diagnostique directement applicable et utile au praticien dans sa pratique quotidienne.

\* \* \* \* \*

5

Le gène IBD1prox, situé dans la région promotrice de IBD1, et dont la séquence partielle est dévoilée dans la présente invention, peut lui aussi avoir un rôle important dans la régulation de l'apoptose cellulaire et du processus inflammatoire, comme suggéré par son expression différentielle dans les cellules 10 matures du système immunitaire. La forte association rapportée dans ce travail entre le marqueur de polymorphisme ctg35ExC (situé dans la région transcrive du gène) et la maladie de Crohn, plaide aussi très fortement en faveur de cette hypothèse.

Les maladies inflammatoires de l'intestin sont des maladies génétiques complexes pour lesquelles, à ce jour, aucun gène de susceptibilité n'avait été 15 identifié avec certitude. L'invention a permis de l'identification du premier gène de susceptibilité à la maladie de Crohn, par une démarche de clonage positionnel (ou génétique reverse). Il s'agit là de la première localisation génétique obtenue par une telle approche pour une maladie génétique complexe, ce qui démontre son utilité et sa faisabilité, au moins dans certains cas dans les maladies génétiques complexes.

20 La présente invention concerne aussi un acide nucléique purifié ou isolé caractérisé en ce qu'il code pour un polypeptide possédant un fragment continu d'au moins 200 acides aminés d'une protéine choisie parmi SEQ ID N° 2 et SEQ ID N° 5.

## Références

Auphan et al. (1995) *Science* 270, 286-90.

Asakawa et al. (1997), *Gene*, 191, 69

5 Becker et al. (1998), *Proc Natl Acad Sci USA*, 95, 9979

Bertin et al. (1999), *J Biol Chem*, 274, 12955

Buckholz, (1993), *Curr. Op. Biotechnology* 4, 538.

Carter, (1993) *Curr. Op. Biotechnology* 3, 533.

Cho et al. (1998), *Proc Natl Acad Sci USA*, 95, 7502.

10 Duck et al. (1990), *Biotechniques*, 9, 142.

Edwards et Aruffo (1993), *Curr. Op. Biotechnology*, 4, 558.

Epstein (1992) *Médecine/Sciences*, 8, 902.

Guatelli et al. (1990), *Proc. Natl. Acad. Sci. USA* 87: 1874.

Hugot et al. (1996), *Nature*, 379, 821.

15 Inohara et al. (1999) *J Biol Chem*, 274, 14560.

Inohara et al. (2000) *J. Biol. Chem.*

Kievitis et al. (1991), *J. Virol. Methods*, 35, 273.

Kim et al., (1996) *Genomics*, 34, 213.

Köhler et Milstein. (1975) *Nature* 256, 495.

20 Kwoh, et al. (1989), *Proc. Natl. Acad. Sci. USA*, 86, 1173.

Landegren et al. (1988) *Science* 241, 1077.

Lander et Kruglyak (1995) *Nat Genet*, 11, 241.

Luckow (1993), *Curr. Op. Biotechnology* 4, 564.

Martin et al. (2000), *Am. J. Hum. Genet.* 67 : 146-54.

25 Matthews et al. (1988), *Anal. Biochem.*, 169, 1-25.

McKay (1999) *Gastroenterol.* 13, 509-516.

Miele et al. (1983), *J. Mol. Biol.*, 171, 281.

Neddeleman et Wunsch (1970) *J. Mol. Biol.* 48 : 443

Ogura et al. (2000), *J. Biol. Chem.*

30 Olins et Lee (1993), *Curr. Op. Biotechnology* 4 : 520.

Perricaudet et al. (1992). *La Recherche* 23 : 471.

Pearson et Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85 : 2444

Poltorak et al. (1998) *Sciences* 282, 2085-8.

Rionyx et al. (1998) *Gastroenterology*, 115: 1062.

Rohmann et al. (1996) *Nature Biotech.* 14 : 1562.

Rolfs, A. et al. (1991), Berlin : Springer-Verlag.

Rouquier et al. (1994), *Anal Biochem* 217, 205.

5 Sambrook et al. (1989) *Molecular cloning : a laboratory manual*. 2<sup>nd</sup> Ed. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Satsangi et al. (1996), *Nat Genet*, 14 : 199.

Schreiber et al. (1998) *Gut* 42, 477-84.

Segev, (1992), Kessler C. Springer Verlag, Berlin, New-York, 197-205.

10 Smith et Waterman (1981) *Ad. App. Math.* 2 : 482

Stewart et Yound (1984), *Solid phase peptides synthesis*, Pierce Chem. Company, Rockford, 111, 2ème éd., (1984).

Spielman et al. (1993) *Am J Hum Genet*, 52, 506.

Sundberg et al. (-1994) *Gastroenterology* 107, 1726-35.

15 Temin, (1986) *Retrovirus vectors for gene transfer*. In Kucherlapati R., ed. *Gene Transfer*, New York, Plenum Press, 149-187.

Tromp et al. (1996) *Am J Hum Genet*, 59 : 1097.

Wahl et al. (1998) *B. J. Clin. Invest.* 101, 1163-74.

Walker (1992), *Nucleic Acids Res.* 20 : 1691.

**Revendications**

1. Acide nucléique purifié ou isolé, caractérisé en ce qu'il comprend une séquence nucléique choisie dans le groupe de séquences suivantes :

5 a) SEQ ID N° 1, SEQ ID N° 3, SEQ ID N° 4 et SEQ ID N° 6 ;  
b) la séquence d'un fragment d'au moins 15 nucléotides consécutifs d'une séquence choisie parmi SEQ ID N° 1, SEQ ID N° 3, SEQ ID N° 4 ou SEQ ID N° 6 ;  
c) une séquence nucléique présentant un pourcentage d'identité d'au moins 80 %, après alignement optimal avec une séquence définie en a) ou b) ;  
10 d) une séquence nucléique s'hybridant dans des conditions de forte stringence avec une séquence nucléique définie en a) ou b) ;  
e) la séquence complémentaire ou la séquence d'ARN correspondant à une séquence telle que définie en a), b), c) ou d).  
15

2. Acide nucléique purifié ou isolé selon la revendication 1, caractérisé en ce qu'il comprend ou est constitué d'une séquence choisie parmi SEQ ID N° 1 et SEQ ID N° 4, la séquence complémentaire ou la séquence d'ARN correspondant à une de ces séquences.  
20

3. Acide nucléique purifié ou isolé caractérisé en ce qu'il code pour un polypeptide possédant un fragment continu d'au moins 200 acides aminés d'une protéine choisie parmi SEQ ID N° 2 et SEQ ID N° 5.  
25

4. Polypeptide isolé caractérisé en ce qu'il comprend un polypeptide choisi parmi :

30 a) un polypeptide correspondant à SEQ ID N° 2 ou SEQ ID N° 5 ;  
b) un polypeptide variant d'un polypeptide de séquence définie en a) ;  
c) un polypeptide homologue à un polypeptide défini en a) ou b), comportant au moins 80 % d'homologie avec ledit polypeptide de a) ;

- d) un fragment d'au moins 15 acides aminés consécutifs d'un polypeptide défini en a), b) ou c) ;
- e) un fragment biologiquement actif d'un polypeptide défini en a), b) ou c).

5

5. Polypeptide selon la revendication 4, caractérisé en ce qu'il est constitué d'une séquence choisie parmi SEQ ID N° 2, SEQ ID N° 5 ou une séquence possédant au moins 80 % d'homologie avec l'une de ces séquences après alignement optimal.

10

6. Vecteur de clonage et/ou d'expression comprenant un acide nucléique selon l'une des revendications 1 à 3 ou codant pour un polypeptide selon l'une des revendications 4 et 5.

15

7. Cellule hôte caractérisée en ce qu'elle est transformée par un vecteur selon la revendication 6.

8. Animal, excepté l'homme, caractérisé en ce qu'il comprend une cellule selon la revendication 7.

20

9. Utilisation d'une séquence d'acide nucléique selon l'une des revendications 1 à 3 en tant que sonde ou amorce, pour la détection et/ou l'amplification de séquences d'acide nucléique.

25

10. Utilisation *in vitro* d'un acide nucléique selon l'une des revendications 1 à 3 comme oligonucléotide sens ou antisens.

11. Utilisation d'une séquence d'acide nucléique selon l'une des revendications 1 à 3 pour la production d'un polypeptide recombinant.

30

12. Procédé d'obtention d'un polypeptide recombinant caractérisé en ce que l'on cultive une cellule selon la revendication 7 dans des conditions permettant l'expression dudit polypeptide et que l'on récupère ledit polypeptide recombinant.

13. Polypeptide recombinant caractérisé en ce qu'il est obtenu par un procédé selon la revendication 12.

5 14. Anticorps monoclonal ou polyclonal caractérisé en ce qu'il lie sélectivement un polypeptide selon l'une des revendications 4, 5 ou 13.

15. Procédé de détection d'un polypeptide selon l'une des revendications 4, 5 ou 13, caractérisé en ce qu'il comprend les étapes suivantes :

10 a) mise en contact d'un échantillon biologique avec un anticorps selon la revendication 14 ;  
b) mise en évidence du complexe antigène-anticorps formé.

16. Trousse de réactifs pour la mise en œuvre d'un procédé selon la 15 revendication 15, caractérisée en ce qu'elle comprend :

a) un anticorps monoclonal ou polyclonal selon la revendication 14 ;  
b) éventuellement des réactifs pour la constitution d'un milieu propice à la réaction immunologique ;  
20 c) les réactifs permettant la détection du complexe antigène-anticorps produit lors de la réaction immunologique.

17. Méthode de diagnostic et/ou d'évaluation pronostique d'une maladie inflammatoire et/ou immune ou d'un cancer caractérisée en ce qu'on détermine à 25 partir d'un prélèvement biologique d'un patient la présence d'au moins une mutation et/ou une altération d'expression du gène correspondant à SEQ ID N° 1, SEQ ID N° 3, SEQ ID N° 4 ou SEQ ID N° 6 par l'analyse de tout ou partie d'une séquence nucléique correspondant audit gène.

30 18. Puce à ADN caractérisée en ce qu'elle contient une séquence nucléique selon l'une des revendications 1 à 3.

19. Puce à protéines caractérisée en ce qu'elle contient un polypeptide selon l'une des revendications 4, 5 ou 13, ou un anticorps selon la revendication 14.

20. Procédé de détection et/ou de dosage d'un acide nucléique selon l'une 5 des revendications 1 à 3 dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mise en contact d'un polynucléotide selon l'une des revendications 1 à 3, marqué ;
- b) détection et/ou dosage de l'hybride formé entre ledit 10 polynucléotide et l'acide nucléique de l'échantillon biologique.

21. Procédé de détection et/ou de dosage d'un acide nucléique selon l'une des revendications 1 à 3 dans un échantillon biologique, caractérisé en ce qu'il comprend une étape d'amplification des acides nucléiques dudit échantillon 15 biologique à l'aide d'amorces choisies parmi les acides nucléiques selon l'une des revendications 1 à 2.

22. Procédé de criblage de composés capables de se fixer à un polypeptide de séquence SEQ ID N° 2 ou SEQ ID N° 5, caractérisé en ce qu'il comprend les 20 étapes de mise en contact d'un polypeptide selon l'une des revendications 4, 5 ou 13, d'une cellule selon la revendication 7, ou d'un mammifère selon la revendication 8, avec un composé candidat et de détection de la formation d'un complexe entre ledit composé candidat et ledit polypeptide.

25 23. Procédé de criblage de composés capables d'interagir *in vitro* ou *in vivo* avec un acide nucléique selon l'une des revendications 1 à 3, caractérisé en ce qu'il comprend les étapes de mise en contact d'un acide nucléique selon l'une des revendications 1 à 3, d'une cellule selon la revendication 7, ou d'un mammifère selon la revendication 8, avec un composé candidat et de détection de la formation 30 d'un complexe entre ledit composé candidat et ledit acide nucléique

24. Composé caractérisé en ce qu'il est choisi parmi

- a) un acide nucléique selon l'une des revendications 1 à 3 ;

- b) un polypeptide selon l'une des revendications 4, 5 ou 13 ;
- c) un vecteur selon la revendication 6 ;
- d) une cellule selon la revendication 7 ; et
- e) un anticorps selon la revendication 14 ;

5                   à titre de médicament.

25. Composé selon la revendication 24, pour la prévention et/ou le traitement d'une maladie inflammatoire et/ou immune ou d'un cancer associé à la présence d'au moins une mutation du gène correspondant à SEQ ID N° 1 ou SEQ  
10 ID N° 4.

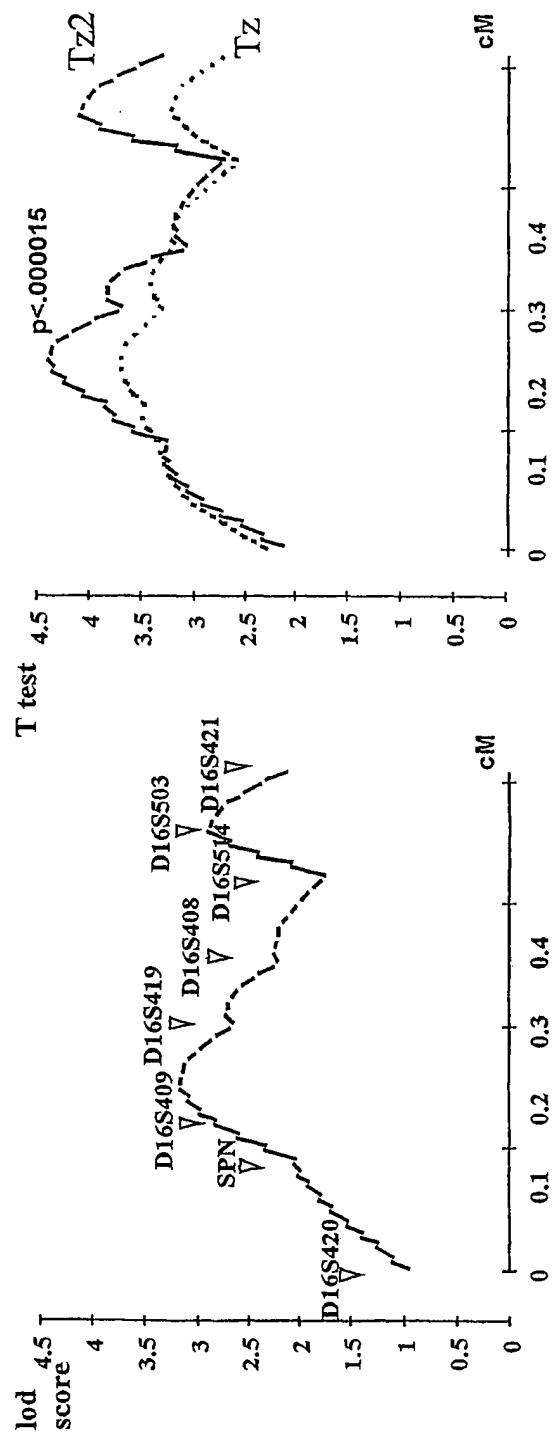


FIG.1

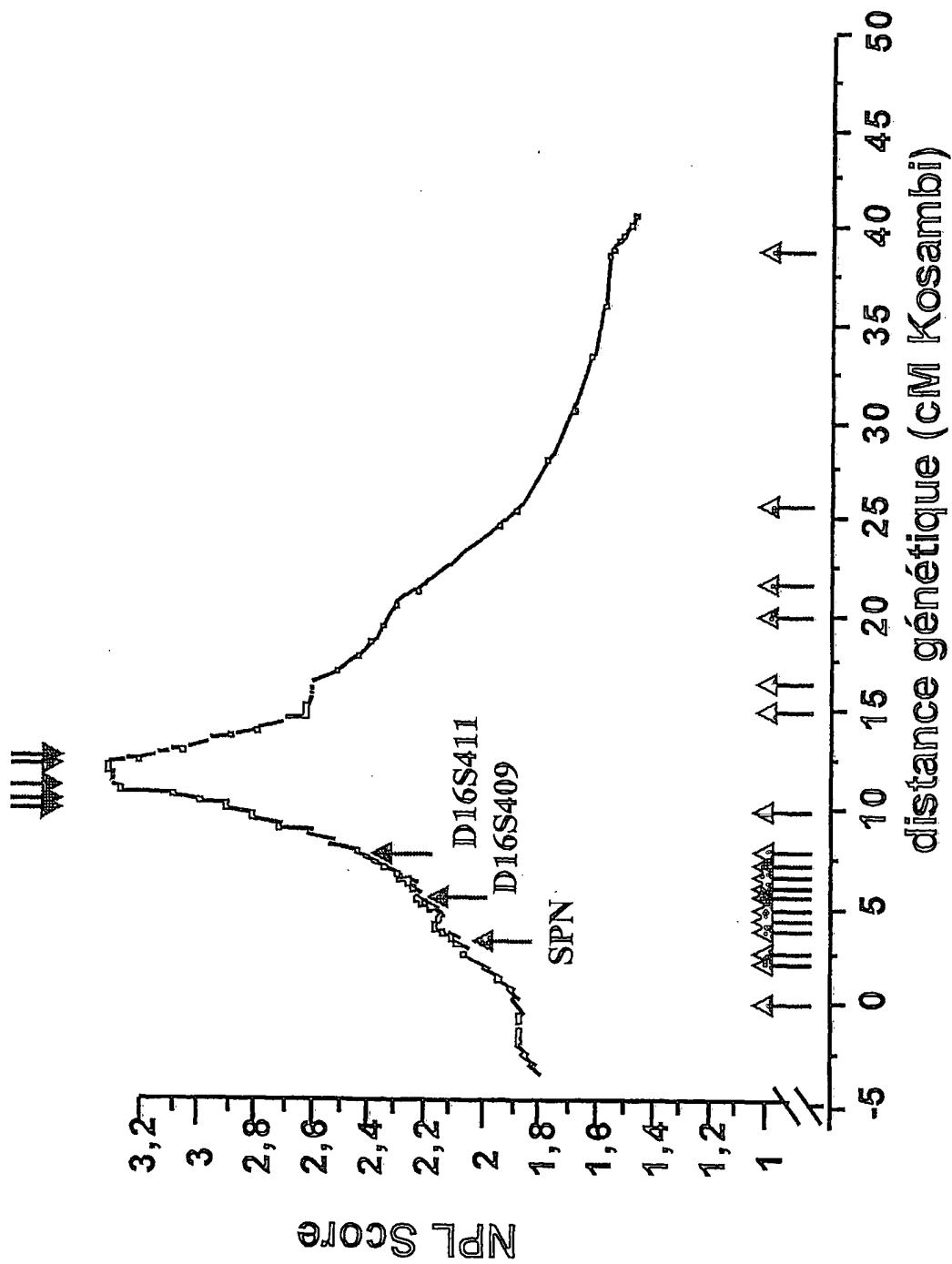


FIG.2

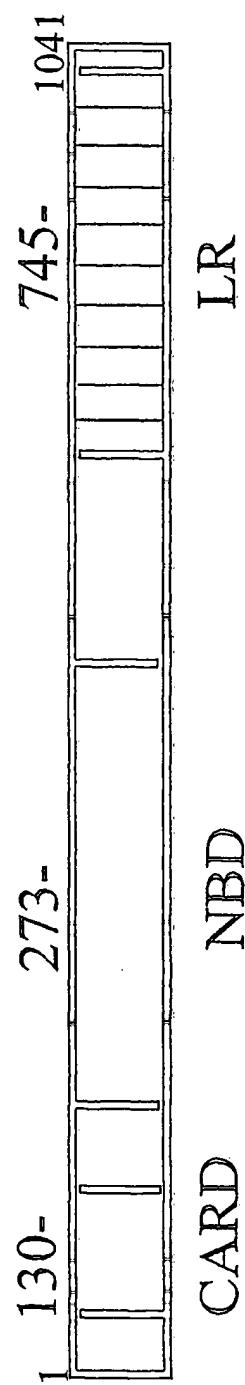


FIG.3

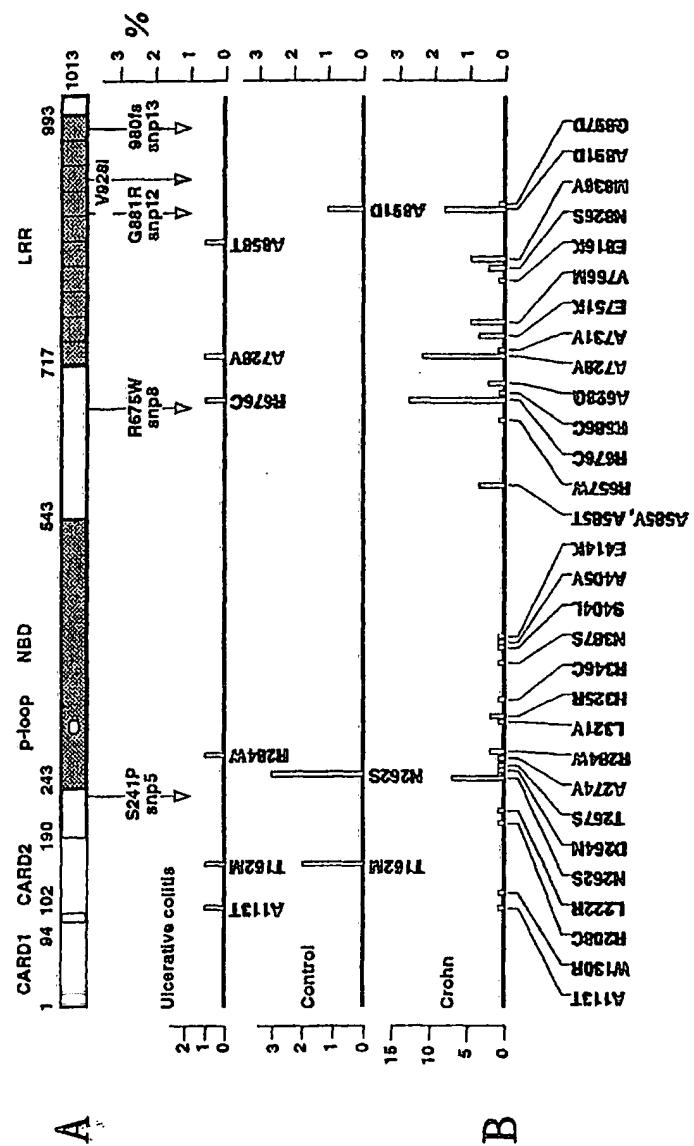


FIG. 4

## LISTE DE SÉQUENCES

&lt;110&gt; Fondation Jean Dausset - CEPH

&lt;120&gt; Gènes impliqués dans les maladies inflammatoires de l'intestin et leur utilisation

&lt;130&gt; D18702

&lt;160&gt; 90

&lt;170&gt; PatentIn Ver. 2.1

&lt;210&gt; 1

&lt;211&gt; 4322

&lt;212&gt; ADN

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)...(3123)

&lt;400&gt; 1

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |    |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| atg | gag | aag | aga | agg | ggt | cta | acc | att | gag | tgc | tgg | ggc | ccc | caa | agt | 48 |
| Met | Glu | Lys | Arg | Arg | Gly | Leu | Thr | Ile | Glu | Cys | Trp | Gly | Pro | Gln | Ser |    |
| 1   | 5   |     |     |     |     | 10  |     |     |     |     |     |     | 15  |     |     |    |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |    |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| ccc | tca | ctg | acc | ttg | ttc | tcc | cca | ggt | tgt | gaa | atg | tgc | tcg | cag | 96 |
| Pro | Ser | Leu | Thr | Leu | Phe | Ser | Ser | Pro | Gly | Cys | Met | Cys | Ser | Gln |    |
| 20  | 25  |     |     |     |     |     |     |     |     |     | 30  |     |     |     |    |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| gag | gct | ttt | cag | gca | cag | agg | agc | cag | ctg | gtc | gag | ctg | ctg | gtc | tca | 144 |
| Glu | Ala | Phe | Gln | Ala | Gln | Arg | Ser | Gln | Leu | Val | Glu | Leu | Leu | Val | Ser |     |
| 35  | 40  |     |     |     |     |     |     |     |     |     | 45  |     |     |     |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| ggg | tcc | ctg | gaa | ggc | ttc | gag | agt | gtc | ctg | gac | tgg | ctg | ctg | tcc | tgg | 192 |  |
| Gly | Ser | Leu | Glu | Gly | Phe | Ser | Val | Glu | Ser | Leu | Asp | Trp | Ieu | Ieu | Ser | Trp |  |
| 50  | 55  |     |     |     |     |     |     |     |     |     | 60  |     |     |     |     |     |  |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| gag | gtc | ctc | tcc | tgg | gag | gac | tac | gag | ggc | ttc | cac | ctc | ctg | ggc | cag | 240 |
| Glu | Val | Leu | Ser | Trp | Glu | Asp | Tyr | Glu | Gly | Phe | His | Ieu | Ieu | Gly | Gln |     |
| 65  | 70  |     |     |     |     |     |     |     |     | 75  |     |     |     | 80  |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| cct | ctc | tcc | cac | ttg | gcc | agg | cgc | ctt | ctg | gac | acc | gtc | tgg | aat | aag | 288 |
| Pro | Leu | Ser | His | Leu | Ala | Arg | Arg | Leu | Leu | Asp | Thr | Val | Trp | Asn | Lys |     |
| 85  | 90  |     |     |     |     |     |     |     |     |     | 95  |     |     |     |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ggt | act | tgg | gcc | tgt | cag | aag | ctc | atc | gct | gac | gcc | caa | gaa | gcc | cag | 336 |
| Gly | Thr | Trp | Ala | Cys | Gln | Lys | Leu | Ile | Ala | Ala | Ala | Gln | Glu | Ala | Gln |     |
| 100 | 105 |     |     |     |     |     |     |     |     |     | 110 |     |     |     |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| gcc | gac | agc | cag | tcc | ccc | aag | ctg | cat | ggc | tgc | tgg | gac | ccc | cac | tcg | 384 |
| Ala | Asp | Ser | Gln | Ser | Pro | Lys | Leu | His | Gly | Cys | Trp | Asp | Pro | His | Ser |     |
| 115 | 120 |     |     |     |     |     |     |     |     |     | 125 |     |     |     |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ctc | cac | cca | gcc | cga | gac | ctg | cag | agt | cac | cgg | cca | gcc | att | gtc | agg | 432 |
| Leu | His | Pro | Ala | Arg | Asp | Leu | Gln | Ser | His | Arg | Pro | Ala | Ile | Val | Arg |     |
| 130 | 135 |     |     |     |     |     |     |     |     | 140 |     |     |     |     |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| agg | ctc | cac | agc | cat | gtg | gag | aac | atg | ctg | gac | ctg | gca | tgg | gag | cgg | 480 |
| Arg | Leu | His | Ser | His | Val | Glu | Asn | Met | Leu | Asp | Leu | Ala | Trp | Glu | Arg |     |

| 145                                                                                                                                | 150 | 155 | 160 |      |
|------------------------------------------------------------------------------------------------------------------------------------|-----|-----|-----|------|
| ggt ttc gtc agc cag tat gaa tgt gat gaa atc agg ttg ccg atc ttc<br>Gly Phe Val Ser Gln Tyr Glu Cys Asp Glu Ile Arg Leu Pro Ile Phe |     |     |     | 528  |
| 165                                                                                                                                |     | 170 |     | 175  |
| aca ccg tcc cag agg gca aga agg ctg ctt gat ctt gcc acg gtg aaa<br>Thr Pro Ser Gln Arg Ala Arg Arg Leu Leu Asp Leu Ala Thr Val Lys |     |     |     | 576  |
| 180                                                                                                                                |     | 185 |     | 190  |
| gcg aat gga ttg gct gcc ttc ctt cta caa cat gtt cag gaa tta cca<br>Ala Asn Gly Leu Ala Ala Phe Leu Leu Gln His Val Gln Glu Leu Pro |     |     |     | 624  |
| 195                                                                                                                                |     | 200 |     | 205  |
| gtc cca ttg gcc ctg cct ttg gaa gct gcc aca tgc aag aag tat atg<br>Val Pro Leu Ala Leu Pro Leu Glu Ala Ala Thr Cys Lys Lys Tyr Met |     |     |     | 672  |
| 210                                                                                                                                |     | 215 |     | 220  |
| gcc aag ctg agg acc acg gtg tct gct cag tct cgc ttc ctc agt acc<br>Ala Lys Leu Arg Thr Val Ser Ala Gln Ser Arg Phe Leu Ser Thr     |     |     |     | 720  |
| 225                                                                                                                                |     | 230 |     | 235  |
| 240                                                                                                                                |     |     |     |      |
| tat gat gga gca gag acg ctc tgc ctg gag gac ata tac aca gag aat<br>Tyr Asp Gly Ala Glu Thr Leu Cys Leu Glu Asp Ile Tyr Thr Glu Asn |     |     |     | 768  |
| 245                                                                                                                                |     | 250 |     | 255  |
| gtc ctg gag gtc tgg gca gat gtg ggc atg gct gga tcc ccg cag aag<br>Val Leu Glu Val Trp Ala Asp Val Gly Met Ala Gly Ser Pro Gln Lys |     |     |     | 816  |
| 260                                                                                                                                |     | 265 |     | 270  |
| agc cca gcc acc ctg ggc ctg gag gag ctc ttc agc acc cct ggc cac<br>Ser Pro Ala Thr Leu Gly Leu Glu Glu Leu Phe Ser Thr Pro Gly His |     |     |     | 864  |
| 275                                                                                                                                |     | 280 |     | 285  |
| ctc aat gac gat gcg gac act gtg ctg gtg ggt gag gcg ggc agt<br>Leu Asn Asp Asp Ala Asp Thr Val Leu Val Val Gly Glu Ala Gly Ser     |     |     |     | 912  |
| 290                                                                                                                                |     | 295 |     | 300  |
| ggc aag agc acg ctc ctg cag cgg ctg cac ttg ctg tgg gct gca ggg<br>Gly Lys Ser Thr Leu Leu Gln Arg Leu His Leu Leu Trp Ala Ala Gly |     |     |     | 960  |
| 305                                                                                                                                |     | 310 |     | 315  |
|                                                                                                                                    |     |     |     | 320  |
| caa gac ttc cag gaa ttt ctc ttt gtc ttc cca ttc agc tgc cgg cag<br>Gln Asp Phe Gln Glu Phe Leu Phe Val Phe Pro Phe Ser Cys Arg Gln |     |     |     | 1008 |
| 325                                                                                                                                |     | 330 |     | 335  |
| ctg cag tgc atg gcc aaa cca ctc tct gtg cgg act cta ctc ttt gag<br>Leu Gln Cys Met Ala Lys Pro Leu Ser Val Arg Thr Leu Leu Phe Glu |     |     |     | 1056 |
| 340                                                                                                                                |     | 345 |     | 350  |
| cac tgc tgt tgg cct gat gtt ggt caa gaa gac atc ttc cag tta ctc<br>His Cys Cys Trp Pro Asp Val Gly Gln Glu Asp Ile Phe Gln Leu Leu |     |     |     | 1104 |
| 355                                                                                                                                |     | 360 |     | 365  |
| ctt gac cac cct gac cgt gtc ctg tta acc ttt gat ggc ttt gac gag<br>Leu Asp His Pro Asp Arg Val Leu Leu Thr Phe Asp Gly Phe Asp Glu |     |     |     | 1152 |
| 370                                                                                                                                |     | 375 |     | 380  |
| ttc aag ttc agg ttc acg gat cgt gaa cgc cac tgc tcc ccg acc gac<br>Phe Lys Phe Arg Phe Thr Asp Arg Glu Arg His Cys Ser Pro Thr Asp |     |     |     | 1200 |
| 385                                                                                                                                |     | 390 |     | 395  |
|                                                                                                                                    |     |     |     | 400  |

|                                                                                                                                                       |      |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| ccc acc tct gtc cag acc ctg ctc aac ctt ctg cag ggc aac ctg<br>Pro Thr Ser Val Gln Thr Leu Leu Phe Asn Leu Leu Gln Gly Asn Leu<br>405 410 415         | 1248 |
| ctg aag aat gcc cgc aag gtg gtg acc agc cgt ccg gcc gct gtg tcg<br>Leu Lys Asn Ala Arg Lys Val Val Thr Ser Arg Pro Ala Ala Val Ser<br>420 425 430     | 1296 |
| gcg ttc ctc agg aag tac atc cgc acc gag ttc aac ctc aag ggc ttc<br>Ala Phe Leu Arg Lys Tyr Ile Arg Thr Glu Phe Asn Leu Lys Gly Phe<br>435 440 445     | 1344 |
| tct gaa cag ggc atc gag ctg tac ctg agg aag cgt cat cat gag ccc<br>Ser Glu Gln Gly Ile Glu Leu Tyr Leu Arg Lys Arg His His Glu Pro<br>450 455 460     | 1392 |
| ggg gtg gcg gac cgc ctc atc cgc ctg ctc caa gag acc tca gcc ctg<br>Gly Val Ala Asp Arg Leu Ile Arg Leu Leu Gln Glu Thr Ser Ala Leu<br>465 470 475 480 | 1440 |
| cac ggt ttg tgc cac ctg cct gtc ttc tca tgg atg gtg tcc aaa tgc<br>His Gly Leu Cys His Leu Pro Val Phe Ser Trp Met Val Ser Lys Cys<br>485 490 495     | 1488 |
| cac cag gaa ctg ttg ctg cag gag ggg ggg tcc cca aag acc act aca<br>His Gln Glu Leu Leu Gln Glu Gly Ser Pro Lys Thr Thr Thr<br>500 505 510             | 1536 |
| gat atg tac ctg ctg att ctg cag cat ttt ctg ctg cat gcc acc ccc<br>Asp Met Tyr Leu Leu Ile Leu Gln His Phe Leu Leu His Ala Thr Pro<br>515 520 525     | 1584 |
| cca gac tca gct tcc caa ggt ctg gga ccc agt ctt ctt cgg ggc cgc<br>Pro Asp Ser Ala Ser Gln Gly Leu Gly Pro Ser Leu Leu Arg Gly Arg<br>530 535 540     | 1632 |
| ctc ccc acc ctc ctg cac ctg ggc aga ctg gct ctg tgg ggc ctg ggc<br>Leu Pro Thr Leu Leu His Leu Gly Arg Leu Ala Leu Trp Gly Leu Gly<br>545 550 555 560 | 1680 |
| atg tgc tgc tac gtg ttc tca gcc cag cag ctc cag gca gca cag gtc<br>Met Cys Cys Tyr Val Phe Ser Ala Gln Gln Leu Gln Ala Ala Gln Val<br>565 570 575     | 1728 |
| agc cct gat gac att tct ctt ggc ttc ctg gtg cgt gcc aaa ggt gtc<br>Ser Pro Asp Asp Ile Ser Leu Gly Phe Leu Val Arg Ala Lys Gly Val<br>580 585 590     | 1776 |
| gtg cca ggg agt acg gcg ccc ctg gaa ttc ctt cac atc act ttc cag<br>Val Pro Gly Ser Thr Ala Pro Leu Glu Phe Leu His Ile Thr Phe Gln<br>595 600 605     | 1824 |
| tgc ttc ttt gcc gcg ttc tac ctg gca ctc agt gct gat gtg cca cca<br>Cys Phe Phe Ala Ala Phe Tyr Leu Ala Leu Ser Ala Asp Val Pro Pro<br>610 615 620     | 1872 |
| gct ttg ctc aga cac ctc ttc aat tgt ggc agg cca ggc aac tca cca<br>Ala Leu Leu Arg His Leu Phe Asn Cys Gly Arg Pro Gly Asn Ser Pro<br>625 630 635 640 | 1920 |

WO 01/72822

PCT/FR01/00935

|                                                                                                                                                       |      |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| atg gcc agg ctc ctg ccc acg atg tgc atc cag gcc tcg gag gga aag<br>Met Ala Arg Leu Leu Pro Thr Met Cys Ile Gln Ala Ser Glu Gly Lys<br>645 650 655     | 1968 |
| gac agc agc gtg gca gct ttg ctg cag aag gcc gag ccg cac aac ctt<br>Asp Ser Ser Val Ala Ala Leu Leu Gln Lys Ala Glu Pro His Asn Leu<br>660 665 670     | 2016 |
| cag atc aca gca gcc ttc ctg gca ggg ctg ttg tcc cgg gag cac tgg<br>Gln Ile Thr Ala Ala Phe Leu Ala Gly Leu Leu Ser Arg Glu His Trp<br>675 680 685     | 2064 |
| ggc ctg ctg gct gag tgc cag aca tct gag aag gcc ctg ctc cgg cgc<br>Gly Leu Leu Ala Glu Cys Gln Thr Ser Glu Lys Ala Leu Leu Arg Arg<br>690 695 700     | 2112 |
| cag gcc tgt gcc cgc tgg tgt ctg gcc cgc agc ctc cgc aag cac ttc<br>Gln Ala Cys Ala Arg Trp Cys Leu Ala Arg Ser Leu Arg Lys His Phe<br>705 710 715 720 | 2160 |
| cac tcc atc ccg cca gct gca ccg ggt gag gcc aag agc gtg cat gcc<br>His Ser Ile Pro Pro Ala Ala Pro Gly Glu Ala Lys Ser Val His Ala<br>725 730 735     | 2208 |
| atg ccc ggg ttc atc tgg ctc atc cgg agc ctg tac gag atg cag gag<br>Met Pro Gly Phe Ile Trp Leu Ile Arg Ser Leu Tyr Glu Met Gln Glu<br>740 745 750     | 2256 |
| gag cgg ctg gct cgg aag gct gca cgt ggc ctg aat gtt ggg cac ctc<br>Glu Arg Leu Ala Arg Lys Ala Ala Arg Gly Leu Asn Val Gly His Leu<br>755 760 765     | 2304 |
| aag ttg aca ttt tgc agt gtg ggc ccc act gag tgg tct gct gcc ctg gcc<br>Lys Leu Thr Phe Cys Ser Val Gly Pro Thr Glu Cys Ala Ala Leu Ala<br>770 775 780 | 2352 |
| ttt gtg ctg cag cac ctt cgg cgg ccc gtg gcc ctg cag ctg gac tac<br>Phe Val Leu Gln His Leu Arg Arg Pro Val Ala Leu Gln Leu Asp Tyr<br>785 790 795 800 | 2400 |
| aac tct gtg ggt gac att ggc gtg gag cag ctg ctg cct tgc ctt ggt<br>Asn Ser Val Gly Asp Ile Gly Val Glu Gln Leu Leu Pro Cys Leu Gly<br>805 810 815     | 2448 |
| gtc tgc aag gct ctg tat ttg cgc gat aac aat atc tca gac cga ggc<br>Val Cys Lys Ala Leu Tyr Leu Arg Asp Asn Asn Ile Ser Asp Arg Gly<br>820 825 830     | 2496 |
| atc tgc aag ctc att gaa tgt gct ctt cac tgc gag caa ttg cag aag<br>Ile Cys Lys Leu Ile Glu Cys Ala Leu His Cys Glu Gln Leu Gln Lys<br>835 840 845     | 2544 |
| tta gct cta ttc aac aac aaa ttg act gac ggc tgg gca cac tcc atg<br>Leu Ala Leu Phe Asn Asn Lys Leu Thr Asp Gly Cys Ala His Ser Met<br>850 855 860     | 2592 |
| gct aag ctc ctt gca tgc agg cag aac ttc ttg gca ttg agg ctg ggg<br>Ala Lys Leu Leu Ala Cys Arg Gln Asn Phe Leu Ala Leu Arg Leu Gly<br>865 870 875 880 | 2640 |
| aat aac tac atc act gcc gcg gga gcc caa gtg ctg gcc gag ggg ctc                                                                                       | 2688 |

|                                                                     |      |      |      |      |
|---------------------------------------------------------------------|------|------|------|------|
| Asn Asn Tyr Ile Thr Ala Ala Gly Ala Gln Val Leu Ala Glu Gly Leu     | 885  | 890  | 895  |      |
| cga ggc aac acc tcc ttg cag ttc ctg gga ttc tgg ggc aac aga gtg     |      |      |      | 2736 |
| Arg Gly Asn Thr Ser Leu Gln Phe Leu Gly Phe Trp Gly Asn Arg Val     | 900  | 905  | 910  |      |
| ggt gac gag ggg gcc cag gcc ctg gct gaa gcc ttg ggt gat cac cag     |      |      |      | 2784 |
| Gly Asp Glu Gly Ala Gln Ala Leu Ala Glu Ala Leu Gly Asp His Gln     | 915  | 920  | 925  |      |
| agc ttg agg tgg ctc agc ctg gtg ggg aac aac att ggc agt gtg ggt     |      |      |      | 2832 |
| Ser Leu Arg Trp Leu Ser Leu Val Gly Asn Asn Ile Gly Ser Val Gly     | 930  | 935  | 940  |      |
| gcc caa gcc ttg gca ctg atg ctg gca aag aac gtc atg cta gaa gaa     |      |      |      | 2880 |
| Ala Gln Ala Leu Ala Leu Met Leu Ala Lys Asn Val Met Leu Glu Glu     | 945  | 950  | 955  | 960  |
| ctc tgc ctg gag gag aac cat ctc cag gat gaa ggt gta tgt tct ctc     |      |      |      | 2928 |
| Leu Cys Leu Glu Glu Asn His Leu Gln Asp Glu Gly Val Cys Ser Leu     | 965  | 970  | 975  |      |
| gca gaa gga ctg aag aaa aat tca agt ttg aaa atc ctg aag ttg tcc     |      |      |      | 2976 |
| Ala Glu Gly Leu Lys Lys Asn Ser Ser Leu Lys Ile Leu Lys Leu Ser     | 980  | 985  | 990  |      |
| aat aac tgc atc acc tac cta ggg gca gaa gcc ctc ctg cag gcc ctt     |      |      |      | 3024 |
| Asn Asn Cys Ile Thr Tyr Leu Gly Ala Glu Ala Leu Leu Gln Ala Leu     | 995  | 1000 | 1005 |      |
| gaa agg aat gac acc atc ctg gaa gtc tgg ctc cga ggg aac act ttc     |      |      |      | 3072 |
| Glu Arg Asn Asp Thr Ile Leu Glu Val Trp Leu Arg Gly Asn Thr Phe     | 1010 | 1015 | 1020 |      |
| tct cta gag gag gtt gac aag ctc ggc tgc agg gac acc aga ctc ttg     |      |      |      | 3120 |
| Ser Leu Glu Glu Val Asp Lys Leu Gly Cys Arg Asp Thr Arg Leu Leu     | 1025 | 1030 | 1035 | 1040 |
| ctt tgaagtctcc gggaggatgt tcgtctcagt ttgtttgtga caggctgtga          |      |      |      | 3173 |
| Leu                                                                 |      |      |      |      |
| gtttggggcc cagaggctgg gtgacatgtg ttggcagcct cttcaaaatg agccctgtcc   |      |      |      | 3233 |
| tgcctaaggc tgaacttgtt ttctggaaac accataggc acctttattc tggcagagga    |      |      |      | 3293 |
| gggagcatca gtgccctcca ggatagactt ttccccagcc tactttgcc attgacttct    |      |      |      | 3353 |
| tcccaagatt caatcccagg atgtacaagg acagcccccc tccatagttat gggactggcc  |      |      |      | 3413 |
| tctgctgatc ctcccaggct tccgtgtggg tcagtgggc ccatggatgt gcttgttaac    |      |      |      | 3473 |
| tgagtgcctt ttggtgaga ggcggggcc acataattca ggaagcagct ttccccatgt     |      |      |      | 3533 |
| ctcgactcat ccatccaggc cattccccgt ctctggttcc tccccctcctc ctggactcct  |      |      |      | 3593 |
| gcacacgctc cttcctctga ggctgaaatt cagaatatta gtgacacctag ctgtatatt   |      |      |      | 3653 |
| tcacttacag caccccaac cctggcaccc agggtggaa gggctacacc tttagcctgcc    |      |      |      | 3713 |
| ctcccttccg gtgtttaaga catttttggaa aggggacacg tgacagccgt ttgttccccca |      |      |      | 3773 |

agacattcta ggtttgcagaag aaaaatatacg ccacactcca gctyggatca catgtggact 3833  
 tttatttcca gtgaaatcag ttactcttca gttaagcctt tggaaacagc tcgactttaa 3893  
 aaagctccaa atgcagctt aaaaaattaa tctggccag aatttcaaac ggccctacta 3953  
 ggcttctgg ttagtgcctgt gaactgaact ctgacaacag acttctgaaa tagacccaca 4013  
 agaggcagtt ccatttcatt tggccagaa tgctttagga tgtacagttt tggattgaaa 4073  
 gtttacagga aaaaaaattaa ggccgttcct tcaaagcaa tgcatttcctg gattattcaa 4133  
 aatgatgtat gttgaagcct ttgtaaatttgc tcaagatgtc tgcaaatgtt attatttaa 4193  
 acattatgtat gtgtgaaaac tggtaatata ttataggta ctttgttta ctgtcttaag 4253  
 tttatactct tatagacaac atggccgtga actttatgct gtaaataatc agagggaaat 4313  
 aaactgttg 4322

<210> 2  
 <211> 1041  
 <212> PRT  
 <213> Homo sapiens

<400> 2  
 Met Glu Lys Arg Arg Gly Leu Thr Ile Glu Cys Trp Gly Pro Gln Ser  
 1 5 10 15  
 Pro Ser Leu Thr Leu Phe Ser Ser Pro Gly Cys Glu Met Cys Ser Gln  
 20 25 30  
 Glu Ala Phe Gln Ala Gln Arg Ser Gln Leu Val Glu Leu Leu Val Ser  
 35 40 45  
 Gly Ser Leu Glu Gly Phe Glu Ser Val Leu Asp Trp Leu Leu Ser Trp  
 50 55 60  
 Glu Val Leu Ser Trp Glu Asp Tyr Glu Gly Phe His Leu Leu Gly Gln  
 65 70 75 80  
 Pro Leu Ser His Leu Ala Arg Arg Leu Leu Asp Thr Val Trp Asn Lys  
 85 90 95  
 Gly Thr Trp Ala Cys Gln Lys Leu Ile Ala Ala Ala Gln Glu Ala Gln  
 100 105 110  
 Ala Asp Ser Gln Ser Pro Lys Leu His Gly Cys Trp Asp Pro His Ser  
 115 120 125  
 Leu His Pro Ala Arg Asp Leu Gln Ser His Arg Pro Ala Ile Val Arg  
 130 135 140  
 Arg Leu His Ser His Val Glu Asn Met Leu Asp Leu Ala Trp Glu Arg  
 145 150 155 160  
 Gly Phe Val Ser Gln Tyr Glu Cys Asp Glu Ile Arg Leu Pro Ile Phe  
 165 170 175

WO 01/72822

PCT/FR01/00935

Thr Pro Ser Gln Arg Ala Arg Arg Leu Leu Asp Leu Ala Thr Val Lys  
180 185 190

Ala Asn Gly Leu Ala Ala Phe Leu Leu Gln His Val Gln Glu Leu Pro  
195 200 205

Val Pro Leu Ala Leu Pro Leu Glu Ala Ala Thr Cys Lys Lys Tyr Met  
210 215 220

Ala Lys Leu Arg Thr Thr Val Ser Ala Gln Ser Arg Phe Leu Ser Thr  
225 230 235 240

Tyr Asp Gly Ala Glu Thr Leu Cys Leu Glu Asp Ile Tyr Thr Glu Asn  
245 250 255

Val Leu Glu Val Trp Ala Asp Val Gly Met Ala Gly Ser Pro Gln Lys  
260 265 270

Ser Pro Ala Thr Leu Gly Leu Glu Glu Leu Phe Ser Thr Pro Gly His  
275 280 285

Leu Asn Asp Asp Ala Asp Thr Val Leu Val Val Gly Glu Ala Gly Ser  
290 295 300

Gly Lys Ser Thr Leu Leu Gln Arg Leu His Leu Leu Trp Ala Ala Gly  
305 310 315 320

Gln Asp Phe Gln Glu Phe Leu Phe Val Phe Pro Phe Ser Cys Arg Gln  
325 330 335

Leu Gln Cys Met Ala Lys Pro Leu Ser Val Arg Thr Leu Leu Phe Glu  
340 345 350

His Cys Cys Trp Pro Asp Val Gly Gln Glu Asp Ile Phe Gln Leu Leu  
355 360 365

Leu Asp His Pro Asp Arg Val Leu Leu Thr Phe Asp Gly Phe Asp Glu  
370 375 380

Phe Lys Phe Arg Phe Thr Asp Arg Glu Arg His Cys Ser Pro Thr Asp  
385 390 395 400

Pro Thr Ser Val Gln Thr Leu Leu Phe Asn Leu Leu Gln Gly Asn Leu  
405 410 415

Leu Lys Asn Ala Arg Lys Val Val Thr Ser Arg Pro Ala Ala Val Ser  
420 425 430

Ala Phe Leu Arg Lys Tyr Ile Arg Thr Glu Phe Asn Leu Lys Gly Phe  
435 440 445

Ser Glu Gln Gly Ile Glu Leu Tyr Leu Arg Lys Arg His His Glu Pro  
450 455 460

Gly Val Ala Asp Arg Leu Ile Arg Leu Leu Gln Glu Thr Ser Ala Leu  
465 470 475 480

His Gly Leu Cys His Leu Pro Val Phe Ser Trp Met Val Ser Lys Cys  
485 490 495

His Gln Glu Leu Leu Leu Gln Glu Gly Ser Pro Lys Thr Thr

500 505 510

Asp Met Tyr Leu Leu Ile Leu Gln His Phe Leu Leu His Ala Thr Pro  
 515 520 525

Pro Asp Ser Ala Ser Gln Gly Leu Gly Pro Ser Leu Leu Arg Gly Arg  
 530 535 540

Leu Pro Thr Leu Leu His Leu Gly Arg Leu Ala Leu Trp Gly Leu Gly  
 545 550 555 560

Met Cys Cys Tyr Val Phe Ser Ala Gln Gln Leu Gln Ala Ala Gln Val  
 565 570 575

Ser Pro Asp Asp Ile Ser Leu Gly Phe Leu Val Arg Ala Lys Gly Val  
 580 585 590

Val Pro Gly Ser Thr Ala Pro Leu Glu Phe Leu His Ile Thr Phe Gln  
 595 600 605

Cys Phe Phe Ala Ala Phe Tyr Leu Ala Leu Ser Ala Asp Val Pro Pro  
 610 615 620

Ala Leu Leu Arg His Leu Phe Asn Cys Gly Arg Pro Gly Asn Ser Pro  
 625 630 635 640

Met Ala Arg Leu Leu Pro Thr Met Cys Ile Gln Ala Ser Glu Gly Lys  
 645 650 655

Asp Ser Ser Val Ala Ala Leu Leu Gln Lys Ala Glu Pro His Asn Leu  
 660 665 670

Gln Ile Thr Ala Ala Phe Leu Ala Gly Leu Leu Ser Arg Glu His Trp  
 675 680 685

Gly Leu Leu Ala Glu Cys Gln Thr Ser Glu Lys Ala Leu Leu Arg Arg  
 690 695 700

Gln Ala Cys Ala Arg Trp Cys Leu Ala Arg Ser Leu Arg Lys His Phe  
 705 710 715 720

His Ser Ile Pro Pro Ala Ala Pro Gly Glu Ala Lys Ser Val His Ala  
 725 730 735

Met Pro Gly Phe Ile Trp Leu Ile Arg Ser Leu Tyr Glu Met Gln Glu  
 740 745 750

Glu Arg Leu Ala Arg Lys Ala Ala Arg Gly Leu Asn Val Gly His Leu  
 755 760 765

Lys Leu Thr Phe Cys Ser Val Gly Pro Thr Glu Cys Ala Ala Leu Ala  
 770 775 780

Phe Val Leu Gln His Leu Arg Arg Pro Val Ala Leu Gln Leu Asp Tyr  
 785 790 795 800

Asn Ser Val Gly Asp Ile Gly Val Glu Gln Leu Leu Pro Cys Leu Gly  
 805 810 815

Val Cys Lys Ala Leu Tyr Leu Arg Asp Asn Asn Ile Ser Asp Arg Gly  
 820 825 830

Ile Cys Lys Leu Ile Glu Cys Ala Leu His Cys Glu Gln Leu Gln Lys  
 835 840 845  
 Leu Ala Leu Phe Asn Asn Lys Leu Thr Asp Gly Cys Ala His Ser Met  
 850 855 860  
 Ala Lys Leu Leu Ala Cys Arg Gln Asn Phe Leu Ala Leu Arg Leu Gly  
 865 870 875 880  
 Asn Asn Tyr Ile Thr Ala Ala Gly Ala Gln Val Leu Ala Glu Gly Leu  
 885 890 895  
 Arg Gly Asn Thr Ser Leu Gln Phe Leu Gly Phe Trp Gly Asn Arg Val  
 900 905 910  
 Gly Asp Glu Gly Ala Gln Ala Leu Ala Glu Ala Leu Gly Asp His Gln  
 915 920 925  
 Ser Leu Arg Trp Leu Ser Leu Val Gly Asn Asn Ile Gly Ser Val Gly  
 930 935 940  
 Ala Gln Ala Leu Ala Leu Met Leu Ala Lys Asn Val Met Leu Glu Glu  
 945 950 955 960  
 Leu Cys Leu Glu Glu Asn His Leu Gln Asp Glu Gly Val Cys Ser Leu  
 965 970 975  
 Ala Glu Gly Leu Lys Lys Asn Ser Ser Leu Lys Ile Leu Lys Leu Ser  
 980 985 990  
 Asn Asn Cys Ile Thr Tyr Leu Gly Ala Glu Ala Leu Leu Gln Ala Leu  
 995 1000 1005  
 Glu Arg Asn Asp Thr Ile Leu Glu Val Trp Leu Arg Gly Asn Thr Phe  
 1010 1015 1020  
 Ser Leu Glu Glu Val Asp Lys Leu Gly Cys Arg Asp Thr Arg Leu Leu  
 1025 1030 1035 1040  
 Leu

<210> 3  
 <211> 37443  
 <212> ADN  
 <213> Homo sapiens  
  
 <220>  
 <221> exon  
 <222> (63)..(106)  
  
 <220>  
 <221> exon  
 <222> (3908)..(4406)  
  
 <220>  
 <221> exon  
 <222> (12307)..(12412)

<220>  
<221> exon  
<222> (15010)..(16825)

<220>  
<221> exon  
<222> (21017)..(21100)

<220>  
<221> exon  
<222> (21321)..(21404)

<220>  
<221> exon  
<222> (24355)..(24438)

<220>  
<221> exon  
<222> (27052)..(27135)

<220>  
<221> exon  
<222> (27730)..(27813)

<220>  
<221> exon  
<222> (29917)..(30000)

<220>  
<221> exon  
<222> (34244)..(34327)

<220>  
<221> exon  
<222> (36123)..(37443)

<400> 3  
tcaccatata actggattt aaagccacaa gacgggtgg gctcatctag ggatggagt 60  
atatggagaa gagaagggtt ctaaccattt agtgctgggg ccccaagtgt taggaaccag 120  
ccaagaagac agaaagagtg aaaatcagag agttggggtg tcctggagga aatgaagaaa 180  
atgccccaaa gaggaaggag ggaacaaata tgaccaatgc ccctggcaga gcaagcaggc 240  
tgagggctga ggatttggca atggggggtc actgggtgaca gtttactgg agctggatgg 300  
ggaacttagag ggaatggggag gggatggggag gacttggggta cagcagtaca ggcaacagac 360  
aaggggggcct gctgtaaagg gaggcataaa atgggattgg acccaaatga agaaggggag 420  
tgtcaagaga gtgtttact ttacaatgg agaatttagag tgcattgtgc actgggtgggg 480  
ggatttgc tcttagggag agaacagttt tagggagggta gaatgcaggta tagctggggg 540  
agggtggggg gcttggccccc agcagagact caggacactt ggaagtgtg gcttccctgg 600  
gcttcccttc ctctcctgtc tgcaagggtt cagtgggtg agatttcagc acttaagcaa 660  
agcatttgc ctggggccca gagaacccgg gctggctgtg gtctcaggaa ggaaggaggt 720  
gtccaggcgc aggccctggc ctgggtttca gggggggccc acgtgggtca ccccttgacc 780  
ctctcttca gcaaggaagt gatccctttct ctacatgggc ctcaccttgg ggaggacaat 840  
ggtgtctttt aagttgttgt aactgaagta gagatcaaaa ggcaatgcag atagactgac 900  
agatttcgcc tgaagagggg aagcccgacc aggttaaaa ggagtaagag gaaggatgtt 960  
aaggacaatt ttaggaaaca gataatgtt gatattttt tctctctt tcccaattta 1020  
aactgaagca ggagaaactg aagcttagaca taatgattaa ctcccaagc tggtgagctt 1080  
cctgagctgg ttagtggaaa cagcactaag gccaggttct cctccccaga tggtaagat 1140  
gagacaggac aatgcctgtc cagagacagg gcctggctga attggccctc aggattctt 1200  
ctgctctgag gtttctggaa gaaggccagg gcagaggtgt ggtgatgttag ctgctggag 1260  
gacagagctc cgagtacgt ggctggggcg ggcctccct tcttgggtgc cacagaagcc 1320  
caacgtcaact agctgggttg tttatggctc acacgttaggc caggctgccc taggcttgg 1380

gtgcgaaggga gggggccctta cttaatgtgt gcctgtcccc tcgtgaatgt gtctcatgtc 1440  
cccagtgggg ttttcagtg agggtcatgg tctccaggat gcacaaggct ttgtgcaga 1500  
attgcttgg a ttgcctagt tctggaaaggc tggtggcca actctggct ccggctttc 1560  
ctttggaaat ttcccttga ggtgggggtg gtagacagat ccaggtcac cagtcctgtg 1620  
ccactggct tttggcatt tgcacaaggc ctacccgcag atgccatgcc tgctccccc 1680  
gcctaattgg ctgtatggg ggaaggggt ggttcagct ctcacgatga ggagggaaaga 1740  
gcaagtgtcc tcctcgaca ttctccgggt aagaggagca ggcattgtcc cgtcccagct 1800  
tgatcctcag ctttcttca tccttggccg cgacatgtc ccaggcctgg ggtcagatgg 1860  
ggagtgtca ctctgttct gggctgttt ctggggagaa tgggtggcg gggtttttc 1920  
cccaggacct gggcagggtc aatgggtggg gccgctgtcg catccctggc tgggtttcc 1980  
acagctgaga accactccag ggcaagccc agagcttatt ctacccttt ttgtcccttc 2040  
ttccccctgtc ctggccacc ccacccttt ggctcctctg cttagatgtg ggcacaagga 2100  
ggagaactcc ttggccttag agaactaccc tagatcctgg ctccagtg cctctgcagg 2160  
gggttacacc ctctctccca agcagccaga cacacaagta acctcattgc ctcatgttcc 2220  
ccatctgacc agcacaggc cccctgtgcc ccagcagcgt tctgagatg tggagttttc 2280  
tcctttgt taccttggct accgtatgag gacggataca gagtgttccc cccacccca 2340  
gcccaggaga tattttatc atgaacatcc cctcagtg tttgtggggg acaatgtctg 2400  
gccaggctca gggatgccag gacgagtaag acccaggctc ccacgtggcc caggcaggaa 2460  
gagagacaca taaaacacca tcaggaaaga gtaaaatcc ccagggccact tggcatctgc 2520  
tcccttgagt gtctggaaat gtccctgatt tataaaaaga agctgacggc cctttttgtt 2580  
gtccatgtct acacccttc actttcg ttcggggca ctgcagcgc ctttgtccac 2640  
agaccccatg acaatcgac aactgacccat gctgagat tttcttggct gtcaggac 2700  
cctgcccaggc cttgaagctc ctggagggtc acttgcctc aaattccca aacgcacacg 2760  
aggtcaactga tgatagcgt ggcagcgtc tgcacgggt gtttcgagg gctgtggagg 2820  
gaggtgaggg ccctaggc a gttgtgtg ggaagtgtt atggggaca aggacccaga 2880  
acgctcgaa acaacttagt ttgcacccgt a ttttctact tcgccttaga caggacccctt 2940  
agagcaatat tctgagtcta ccccttggag tagcagtg caaaacacac agcacggct 3000  
tggggccccc gtggggaaacc caaatgtaa agtttagagac atgcattccg gactcataca 3060  
tggctcggtg tgaatccctg actctgcctg tctagctgtg acacatgt a caaatca 3120  
agcttctgg tgcctcagtg tcttctctg tagaatgggt agatcatagg cactacttca 3180  
gagtgctgg gagggttcag tgaattctg caggagagca ctttagatgg cacttgggt 3240  
gtagtttag ctaattaaat attagccgtt actgaaactg ctgtacgtt aatccagcca 3300  
gcatgaaaga gcccctctca ccctgttccg aagagaatga attccctgat ttttggaaag 3360  
atctctctct ctctctctgt cttttttttt tttttttag aacggctt gctctcttc 3420  
ccaggctgga ggcataatggt gccatcttgg ctcaactgaa cctctgtccaa 3480  
gtgattctcc tgcctcagcc tcctgatgt ctgggattac aggccgtc caccacgcct 3540  
ggctaatttt ttttttttta gtagagacag ctggccatcc ttttttttttgc gctggcttag 3600  
cgctctgtat ctaaagtgc ctggggagat ctcttgc ttaatattacc tcaaggcttt 3660  
ttaaacgttt taagccggag accaagcatg gatatgggat tttagggct tggatattaa 3720  
cttgggtct tcaaactctg tggaaacccctt aggtgttct tgcctctctt gggctcaat 3780  
tttcacatct atatgggg gacgttggat tggtaatgt ctgaggctag aaccatggcc 3840  
aactcggtt ctgtggggc tgaacttgc tggcccttcc tgaccaccct gcatctggct 3900  
tctggagaag tccctcactg accttggctt cttcccccagg ttgtgaaatg tgctcgcagg 3960  
aggcttttca ggcacagagg agccagctgg tcgactgtctg ggttcaggg tccctggaaag 4020  
gcttcagagag tgcctggac tgctgtctg cttggggagg ctctctctgg gaggactacg 4080  
agggtttcca ctcctgggc cagctctct cccacttgc caggccctt ctggacaccg 4140  
tctggataaa ggtacttgg gctgtcaga agctcatacg ggcctccca gaagcccaagg 4200  
ccgacagcca gttccccaag ctgcatggct gctggaccc ccactcgct caccacgcct 4260  
gagacctgc gacttcccg ccagccattg tcaggagat ccacacccat gttggagaaca 4320  
tgctggaccc ggcataggg ggggggttcg tgcctgtat ttttttttttgc gatcttggct 4380  
tgccgatctt cacaccgtcc cagagggtga ggcactctgtg tggtgcatca caggttctc 4440  
aggaaaagggg tgcttagtca ccaagactga tttgtctca tgaatgtc cttgtgggtt 4500  
acttggtccg tgggatttcc cttaaaagg tagccagca gttaaaattt gctcttgc 4560  
cttggcagga aacatacaac tctttcttcc ttcttttctt ttttttttttgc cactctgtt 4620  
ccctggctag aatgcagtgg cacaatcata gctcaatgtg gccttgaatt cctgcgtca 4680  
agtgtatctt tggccctttaga gtagctgggat ttttttttttgc ttttttttttgc ttttttttttgc 4740  
atttttttt ttttttttag agatgggggtg tgcctatgtt gcccaggctg gctccagct 4800  
cctggcttta agcaatccctc ccgccttggc ttttttttttgc ttttttttttgc ttttttttttgc 4860  
ccactttggcc tggccaaacag aacacttctg cccaggatcc ttttttttttgc ttttttttttgc 4920  
cagattctgg agccagaatg gtcaggctc ttttttttttgc ttttttttttgc ttttttttttgc 4980  
ctatggagcc ttcacccgtcc ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 5040



cctacatgggt cttctctgcct tggctctccctg tattcctgaat cctatggccct cttctccctt 8760  
ggtttactac attttgcgt tag accgtatcc ccaagtcattt cctttagaaatg aatgtatgaa 8820  
agttaaaatt tctgaggctc cacaatgtctt aaagttccct cataactggat tgatgtttt 8880  
gctgggtata aaattctggg ctggccatca ttttccttca gaattttgtat tgcatattt 8940  
cattatcctc tctttcaat attgcttctca agaattccaa aaccctttttt tttttttctt 9000  
tttgagacag tgcgtcaactc tgcacccag gctggatgc agtagtgc tctcagctca 9060  
ctgcaacccctc cacccctgg gtttaagcgta ttcttcttcc tcagcctctt gaggcgtgg 9120  
gattacagge acccaccacc acacccttta gtagagatgg ggttttgtca tggggccag 9180  
gctggcttg aacttctgac tttaggtgtat ctgcctactt cggccctccca aagtgcgtgg 9240  
attaaaggcg tgagccacca caccaggcc cccaaaccat tttaaaactc tttctggaaag 9300  
ctttaaaat tttcttttag tccccagaat tttaaaattt caattatgtt ctttgggtt 9360  
cttccattat attagtccacc caagaggtac ttcaatctg gaaacttctc tttttttttt 9420  
gaaatgtct tgattagttt acagggtattt tttccctctc cattttatct tttctctttt 9480  
catgaaacta ctattaattt aatgttagaa ttcccttgcact gatcattttt tttcttctca 9540  
tttccatct ctgtgtctt ttgcctactt ttctatgtat agtcacagct tttttttttt 9600  
acttctgagt ttttcatttt tgatgtcatg atttttaattt gcaagaggta ggtttgactg 9660  
attttttttt tgtagatctt actttttgttt tatggatgca acatcttctt tgacttaagg 9720  
atcataagat aggtgggtt tttttttgtt tgtagtactg ttttccaccc tttttttttt 9780  
ttttctacaa gtttcttcc ctttcccccc tttttggctt ctatctccca cattagatgc 9840  
tttctctggg ctcatgatac tttttggttt ttttctcaa gattgacagg taggacttta 9900  
aaacttggg agcatgcggg tgaaacttgc tttactgtt gatattttgg 9960  
agattgcacag tgtttatatc tttagatctc acctccctggg ttgatcaagt tatctgagta 10020  
caccacagac ctttgcctg gggataaaacc agaaaatctgt ttcaagaaacc actttgattc 10080  
agtcttcctt gttttagtca ttcccttgc tttccggaggt ccgtcatgtt gatcattcca 10140  
gagccctta cagatccctag ggtacacact gcatgggtt caactttctt gttttgggtt 10200  
taagatttgg ctttcaggag tttccctcagt ccgttactat tcattcaatc agcaagtctt 10260  
tgagcacctg attttgtccca gacatttttc taggtgttag ggataccca gtaacaaaaa 10320  
cagacaaaaa tctttgtctt gaaatacacac acactccagt caggggagag ggacaataag 10380  
ccaaaggaaag gaaattacag ctgtgtctag aaggtgataa gtgtgttaga aagtaagtaa 10440  
agttgggttgg ggagttgaga gtttgggaaag gggataaaatg atggcaattt taaatagagt 10500  
agtcagagtt ctcactttaga aggtgaaatg caagtaaaga cttaaggag gacagggaaat 10560  
tagccacatg gatggcttag ggaaggcttc caagctgaga ggacagccag agccaaggcc 10620  
cagggcagg agcatacttgc tttttttttt gaaacaggag gccaggatgc tgagtggagt 10680  
aagagggggc atgaaaaggag aaacttgggtt ccacgtgggtt cttagacaggat ttttttttgc 10740  
gttttgggccc ctgaaggtaa ctattggact tggacttta ctctgaggaa atagggacgc 10800  
tattgggacg tttgtacagg agcaatgttgc cttttttttt gttttttttt gattagactc 10860  
tggctgtggc attaaggcttgc ggctgtgggg gcaaggaaacag aagcaggggg accagttttt 10920  
cagccctgtgc agtttccag ataaggcaggg attgtgtctt ggaggaggat ggtatagagg 10980  
aggtgacaag aaatgactct atgtctggta tttttttttt ggcacacatg ggcattttttt 11040  
caactagagac ctggctggc cacaatggat ttccataatgc acataatata catcagattt 11100  
caaagactta atatgaaaaa aaaaattttaa cggggccccgg gaattttttt tttttttttt 11160  
ttttttgaga cccagtcttgc ctctgtcacc caggctggag tgcagtgggt tgatctccgc 11220  
tcactgcac ctccgcctcc caggttcaag tttttttttt tgattcttgc ccttgcattttt 11280  
tgggactaca ggcacccgtt accacgcctg gctaattttt tttttttttt tgatgtatgg 11340  
ggtttccacca tggctgtccag gctggcttgc aactccggac cttaggggat ctaccgcct 11400  
tggcctccca aatttgcgttggg attacaggca tgagccacca tgctcagccca tatcttgc 11460  
ttttctacat ggattacatg ttgaaatggt aatgttttttgg ctattgttgc tttttttttt 11520  
tatatgatta aagttgattt catctattttt ttttaactt aaaaatatg tttttttttt 11580  
gatttgaat tccacatgcg gtttgcattt gtgacccttgc tttttttttt tggaacagt 11640  
gcccttttttgg gacatgttgc tggatgggaa gtcacacagga tttttttttt tttttttttt 11700  
aggttcaag ggtgtacttca agacttcggg gcaaggaccc tggaaagaaag gggtaatata 11760  
tagccaaatg gggaaaggctt gtcgggtttgg caggtgcattt ggcaggatgg gttttttttt 11820  
ttgaatatgt tggaggtgtt tatgaaactt tttttttttt tttttttttt tggaacagt 11880  
tgcgtcaatc cagggttcag ggagacaggat caggctggag atgaagatgt gggagcttgc 11940  
ggagagatgg tatttcaataa ttcacatccat gagacttgc gaaatccattt ctcttccaaa 12000  
tgattttacat cctgcagaat cttttccctt atctttgttgc gttttttttt tttttttttt 12060  
tcattttttt ttcagttattt cactgttttgc tttttttttt tttttttttt tggaacagt 12120  
catgcgtca attcaccatc caacactgttgc tttttttttt tttttttttt tggaacagt 12180  
gttgggggggg tgaccccttta ttctggatgg aagagagatg cttatgttgc tttttttttt 12240  
agtaagcctt cccacatttgc tccatcagcc tttttttttt tttttttttt tggaacagt 12300  
ctgttagggcaaa gaaggctgttgc tgatcttgc tttttttttt tttttttttt tggaacagt 12360

cttctacaac atgttcagga attaccagtc ccattggccc tgcctttggga aggttaggtgt 12420  
 atgttctcg ttaatcagaa agggaaaggc agtcagtgc gatccatggtaaagagcaga 12480  
 acacacctcg gttAACatcc catatgcgg cagtagatcc tccctatgac tcaatttcct 12540  
 tgTTTtaagg cttagcaccac cccgtctcat tgggattttg ggagcattaa aaggacaaaa 12600  
 gCGTgtatgg ttagcttata gctttcatia tctccacac agtataactga caattgggct 12660  
 accatataatt gagggtcaac taaagggtt acttaccatc caaactctca ttatctgtac 12720  
 cgaaaagata tggacacatg ttttaggtt gggctggat ctcttgatct ctgaaattta 12780  
 gcagctcaca atgggaaact caagaacca gttggatctag agactctggatccctcagt 12840  
 gcccagggtc accacccaaa ctcaggaaca ggaggggctt ggacccgacc acttgaacat 12900  
 accaggcata ctggccagggtg ctttatggac aatgtctacc ctttgcacca accctgagaa 12960  
 gtaggtggtg ttttttcca ctttatagat gtggaaactg ggcaggggagg ttaagtgac 13020  
 agggagggga agatgggtct gattgtaaat tgcctccacc tacactttct cttttcttgg 13080  
 gagaagaaat gtcagttgt aagagagat gcaagctgg cacttttag ggcttggcc 13140  
 tacaccactg tagggaaagc tcattggac tgaagcccc tgagctgtgt gtgggtctgg 13200  
 cagatgggtc tattcacccctg gactgtgtcc tctggcagc aagcaagcccttgggggg 13260  
 tggctggaag tctgtgcctg gcactcgcga gtgcaccgtc tcattgaaga acaggatcta 13320  
 aacatcaatg cggccacagca gggtgcgcgg cacggagtgc agggccctggg ttggcccttg 13380  
 gttgaggtt gctgtgaca tcatcaagca cagctgtca ctgtaagacc agggcagggt 13440  
 gcaagattcc ccacacttct aaaggtgaca attgggttat ttatttctct ataaaatgac 13500  
 atttttttt tctggagaat ttttagtatca ttgggtatga ctggaaaacc tgcatacgaa 13560  
 atcaggctgg aagaggaaga tatatatctg atatgtactg gagaggaaga tatctatctt 13620  
 atggctcaag ttcaagggtc ctggtatatt cagagggcag aaagctcagc aataatcatc 13680  
 aactctggga acagaggtga cataaaacaca gggcgcccc tttgtgtgac tgcagatagt 13740  
 catcagttagtgc ttcagagatc tatgaaaatt acttgcgtat ttttgggttg aaaatagtg 13800  
 gccagtggtt ggtggggggc agtgaggctg tgatggcggg ggaccatgcc aagctccatc 13860  
 cagcctggga cgctaaacca gcacttcccc atttccctgaa aggggaacta aactctgaca 13920  
 caggaatgg tttgtcttgc ttacttttagt gatgagaaag gaagagcact ggccttccaa 13980  
 acacaccccg tgcataaaaa ctctccctgc atgggggtca tggggaggat ggggaagtgg 14040  
 aggccaggatc acagacttgc ttccggatgc tcagctggg caccgggtg accccggggc 14100  
 ctcccttgc taggtccacc cagataatc aggttccatct ccccatctcg aagtttaact 14160  
 ttatcacatc tcaagttcc ttttgcacg taaggtaaca tattcacagg ttctgagaat 14220  
 ccggacatgg acatcttgc gggctatttgc ttgtgcctatcatatccatg aataataatg 14280  
 ataataagca ccatttttg agagtttgcg atgtcagata ttcttttaaa ctgtatttt 14340  
 tctcgtgtcc tcctgaaaaaa atccctccatg gtgttatatttgc tcccccatttt tacagatgag 14400  
 agaactgagg cccagaaagg ctaaatggct tgcccaatgt tatgggtggac ccaggtttc 14460  
 aaactcaggt gtgtctggct tcagagactg ggctccatgc cccttaagcc ttgttttccc 14520  
 ctttagaaaaa agtcacctga ggctgagttg tgaaggattatccaaatggc caccggcca 14580  
 ctatggcagg acagatatacgaatacaggcttccatgc cagccccagag ccccttccc 14640  
 tcatctagaa ctccctctgg tgcgtatatt gataacggca gtcactgtatg tcttttgc 14700  
 acttacttttgc ttttgcacg ttacactgtg ctaagacttgc gacataggatc atcttagttg 14760  
 atccgtgtaa aactctgtgaa ggttagtgc accatcttc cccacccatc gaggtggaaa 14820  
 ctgaggggtt ggaagttcc ttgcgtgtcc tcaaaatgc cagcttgcg atggaggagc 14880  
 caggatggc gcccgtggc tctccatcc cttcagttat gtcagcggtcc cccgcagcag 14940  
 cccattttgc ggttaggtcc cgtcttcacc atgggtccac ctccatctgc ctcttcttct 15000  
 gccttccatgc tgccacatgc aagaagtata tggccaaatgc gaggaccacg gtgtctgtc 15060  
 agtctcgctt cctcgttacc tatgtatggcagcactgtctgc gacatataca 15120  
 cagagaatgt cctggagggtc tgggcagatg tgggcattgc tgatcccccg cagaagagcc 15180  
 cagccaccccttggcggc gggcttgcggc gagctttca gcacccctgg ccacccatc gacgatgggg 15240  
 acactgtgtctt ggtgggtgggt gaggccggca gtggcaagag caccgtctgc cagccggctgc 15300  
 acttgcgtgtc ggctgcaggc caagacttcc aggaatttcttgc ctttgccttc ccattcagct 15360  
 gcccggcagct gcagtgcgtt gccaaacccac tctctgtgc gactctactc ttgagact 15420  
 gctgtttggcc tggatgttgcgtaaagacatccatgc tcttccatgc actccttgc caccctgacc 15480  
 gtgtccctgtt aacccatgc ggtttgcgc agtcaatgttgc ctttgcctgc gatcgtgaac 15540  
 gcccactgtc cccgaccgc cccacccatgc tccagaccct gctttcaac cttctgcagg 15600  
 gcaaccctgtt gaaatgcg cgcagggtt tgaccaggcc tccggccgt gtgtcggt 15660  
 tcctcaaggaa gtacatccgc accgaggatca acctcaagggttgc ctttgcctgc caggccatgc 15720  
 agctgtacccatgttgcgatc catcatgcg cccgggtggc ggacccgcctc atccgcctgc 15780  
 tccaaatgcacccatgc caccgttgcgatc ggcacccatgc ttttgcctgc tggatgttgc 15840  
 cccaaatgcacccatgc caccgttgcgatc ggcacccatgc ttttgcctgc tggatgttgc 15900  
 ttttgcctgc tggatgttgcgatc ctttgcctgc ttttgcctgc tggatgttgc 15960  
 aaggcttgcggg acccaggatctt cttcgccggcc gcctcccccac cttctgcac ctggcagac 16020

tggctctgtg gggcctggc atgtgctgct acgtgttctc agcccagcag ctccaggcag 16080  
 cacaggctcg ccctgatgac atttctcttgc gcttcttgc gctgtccaaa ggtgtcg 16140  
 cagggagtag cgcgccttc gaattccctc acatcactt ccagtcttc tttccgcgt 16200  
 tctacctggc actcagtgc gatgtgccac cagcttgc cagacaccc ttcattgt 16260  
 gcagggcagg caactcacca atggccaggc tccgtccac gatgtgcata cagggctcgg 16320  
 agggaaagga cagcagcgtg gcagcttgc tgcagaaggc cgagccgcac aacccatcaga 16380  
 tcacagcgc cttcctggc gggctgtgt cccgggagca ctggggctgt ctggctgagt 16440  
 gccagacatc tgagaaggcc ctgctctgc gccaggctg tgcccgctgg tgcctggccc 16500  
 gcagctccg caagcacttc cactccatcc cgccagctgc accgggtgag gccaagagcq 16560  
 tgcattccat gcccgggttc atctggctca tccggagcct gtacgagatg caggaggagc 16620  
 ggctggctcg gaaggctgca cgtggctga atgtggcga cctcaagttt acattttgc 16680  
 gtgtggccc cactgagtgt gtcggctgg cctttgtgt gcagcacctt cggcggcccg 16740  
 tggccctgc gctggactac aactctgtgg gtgacattgg cgtggagcag ctgctgcctt 16800  
 gccttgggtg ctgcaaggct ctgttagtgg tggtactggg cattgtgtt caggtatggg 16860  
 ggagcaccat caaggcttaag tggggagca ccgagctggg ctctagaagt ctggggccag 16920  
 cttccctct gccaccctgc tttgcaacac tgccagatc ccttcccttc tggccttaa 16980  
 tttcaatatg tgcattgtgac agccacactt tattgactgg cctatgtgct gggctgg 17040  
 ctatgtttc cggaaatgacc tcattctaaatc tctacaacca ccctgggggg taggcaggaa 17100  
 tgtttatttc tccattatcc ttgacttgc gctcagagaa gtgaagtaac ttgtccagga 17160  
 aatggcagag ctgggggttca caaattgcattt cattctgtt acagggtttc tgcctccac 17220  
 cagtctatgg atacacttca gaggctccct gaaaaccttg aggtcacttg cagaaagttt 17280  
 tgcattgtat gtgcctgtat caggaacaac accaaatcag aggtcacttg tgcccatca 17340  
 gagacttaa cacccttcaacc agatggaaat ttcaggagcc aagaaataga aagtggctgc 17400  
 agggttacaa ctactgttgg attcctgagg tagcacagtg tccaaacacagg atttcagcac 17460  
 taccctgtt gcttagagcc ccagccaaatc agtgcagggtt ttgcctttt gagaatctgt 17520  
 gcccctgaac tggggggctt cttccacat cttggggca ggcaggcagc gagggtgtgc 17580  
 ctaggcctgc ggtatcagcat ggcacagatt ccccaacat cttccagctt gaaaggggat 17640  
 tgccctgtt ctatattagaa cctatagaa agcagaagtt cttagattgaa gttaaaattt 17700  
 attccctggc tccaggggctt ttggctaca cctggatgac cttattgtac cctaagcatg 17760  
 ggacaaacca cttccgtgaga gtattaggtt ggtatacatc ttctctgggg gcaaagcaac 17820  
 aagatttattt tttcatcatg gaccaaacac atggatcccc actagaaact gtgttagt 17880  
 ttttggtaac cctgacatag ggaccatgtt cttaggtt aacgataata acaacataat 17940  
 acataacata tatacgcaat atatatatgtt attatgtca atgaatgtaa atatgattat 18000  
 acccatcatg gtctggagg aaacagatgac cacactaa atgggtgttt tgaggagat 18060  
 ttgaaaacca gattgtttac aagccatggg caggagttttagt gaagagtgg aggggtgg 18120  
 cagggggctg gggtagtaa cagctggggg agggtagact tgaagggggg agggggaggga 18180  
 gactaattag ctggggggaa ggtatggaga cggctgcctg agcttctgc aagtggaaaga 18240  
 atactgcttg gcccataactc ctcaccccaa ctcttgcgt tgccagcgc cttccacca 18300  
 ctggaccat cagggaggcc gagtgggtg tctgtggag tagtccccag gcatcagcct 18360  
 cccaggagcc agggacgggtt agagaagggg gagagtggat ctggccagcc aatggaaaa 18420  
 cagccagcac caaaactctat ttcccttagga gggaggatca tgatactttt agtggaaatt 18480  
 tggaaacctg tctgtggag caatttccct gatagaaata agaatgtgc ttttccctgg 18540  
 tagtagactc agtttttacc ccaagaggcc aggcattactt ggcctgtgt atcctcatag 18600  
 gtcagttccat ctctggaaattt cttgaatggc tcatccatcc ttgatttaggg atgtcccgt 18660  
 gattaccagg gtgtgcgaaa gggctctggg aaacctgtgg gtcgtctctt gtgttcagag 18720  
 aaagggtgagg gtggcctgtt tctagctcat ggtgcgtcaga ctgtgggtgt taaaggcact 18780  
 cgtggcaatg cagattccctg ggcctgcctc tagtgcattcc cattcgttagt gtttgggtg 18840  
 gggcccaagga aacttatatttttccacagac accccctgggtt attctgatac aagtggctc 18900  
 gccctggggag aactactgtgt ctgcagcaac cagcttgcgtt ttccatttagc aattactgtc 18960  
 cttgagcggag ttttactgtctt cttcaccttta cacacactaa aactgcctaa ggcgttaggg 19020  
 aggggaagca accatgaggt tgctgtgttgcactgtgt tggtgtgttg tgggtgtgt 19080  
 tgcattgtgtg tgcattgtgag agagagagat attgagaaag agaggaaggg aggaaggggg 19140  
 agggcacagg ctccctctcc acagtgcctt cctgccttc tcccacttgc agcgtttca 19200  
 tgccaaactgaa atccctcagc ctcttaggaaa cccttatatac acagtgccttcc tatatagtt 19260  
 tcttttagact ctgcgtctct cagactctgt agtgcattgtt taaaaggattt tatgttaccc 19320  
 acagagagag agcacgcacc accatgtaaa catggaaacct aagtttcaca aaatgactt 19380  
 gctttatgaa ctctggacaca ctctgtcttc ttctgttgc ttttatttcc atttttagaaa 19440  
 tgctgtcgat gacccatggggaa atgatttgc tgcacgttgc cctgcgttgc gaaaaatcac 19500  
 tgcactacag aagtggccat aagaggccctt gagggagaag ctgcacaatg tcatggtaa 19560  
 gagtgggggtt tggagccaaag cgccttaggc tcaaagcctt tatgtgcctg acacacctgg 19620  
 caaagtcaact tgcgttgc tgcctcgtt ttcttctca cgaatgtca taataatgtt 19680

tcccatattca ctggcttgg tggaggatga aatagtta ttattgagaa gtggtaagg 19740  
tagtgcacat tgctagcgat catgattcta ggtgactttt actgtgtacc ggggtgcac 19800  
aaggcttat gtgcacagcc tggtaggct gataatacta ttgttccctc ttttttttt 19860  
ttggaaacgg agtctcgttc tggtggccag gctggggta cagtggcaca atctcggtc 19920  
atgcaatctc tgccctccgg gttcacgcca ttctctgc tcagcctccc aagtagctgg 19980  
gactacaggg gcctgcccacc acgcccggct aattttttt tatttttggg agcgacaggg 20040  
tttcactgtg ttaaccagga tggtctcgat ctccgtaccc cgtgatccgc cccgctccgc 20100  
ctcccaaaatg gctgggatta cagggcgtgag ccaccgtgc cggcctgttc cctttttat 20160  
agatgaagag accagcaat aacttagtaag tcgctgatca ggtacacaat atccagctga 20220  
ggcactccag agcctgagct gttaccatt cagtcaggcc ctcccaagtt tgccctaaaga 20280  
taaagaatca tgcacagt tggtaaaata tacagattcc tggggccac cccgcagata 20340  
cttgattgac agtccaggg tatgggcctg agaatctgtc ttttagggaa gctttagat 20400  
gatgttgtga tcaggtgagt tttggaaatg gtgcggcaag aggagtgcc gacagggtt 20460  
gctcgccagg gactagccctg ttggagtgg tccattgggg ttaaggactg ggcagcagg 20520  
cctcaactaac cacagcctat atgcctgtt ctgaagttt ggcactctc atccagctgg 20580  
tctactgtct gctgacctag atgatggtaa attgtccccca ggggtagcct gtctagttca 20640  
ggctgcaccc ttgcataata tcagctccct tccaccatca tccctttgt gaggctgctg 20700  
tgattatcat gttccctttt cagagatgga aacattgcct caaattagct ctgtcatttc 20760  
ctaaggattc cagggttctt tagtagggg tctggatcc acgtctccgg ccatccccat 20820  
catagtgcac cacgtcacct ccctggccag ggaccgtggg gtcctccactt tttgggttg 20880  
ctccatctat gcagggtttc tggaaatg agatgtggc acttcaggga tgaatgaaag 20940  
tcttttggg ggattttgtat attttttct tgcctacta gtccttattt caaatgtatt 21000  
tattttgtct cttagtttgc gacaaatggc atatctcaga ccgaggcatc tgcaagctca 21060  
ttgaatgtgc tcttcactgc gagcaattgc agaaggtagc gtaagtcagc ctggctgtg 21120  
gacaatgggc tccaaatgtcc tcggatccatc cccaggtcgt gcaagctggg aagctgtgag 21180  
tgatgggctg gggcaggggc gggaaaggc accctgggat tggactccatc atgatgggg 21240  
tctctctgg aactgaacag tctattcaac aacaattga ctgacggctg tgcacactcc 21360  
atggctaaatc tccctgcattt ttttgcattt ttcttgcattt ttcttgcattt ccaggttttc 21420  
cttattccct gaaactatt tggcactcc acactggct aacatgggt ttttgcattt ttcttgcattt 21480  
tggcactcc acactggct aactgaacag tctattcaac aacaattga ctgacggctg tgcacactcc 21540  
aaggcagggt tcttaatgc acttcctgtc ttttgcattt ttcttgcattt ttcttgcattt ccaggttttc 21600  
acttcctgtc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 21660  
tttctgtgtc taagagtgtc acagcttttct ggggttactg agttccacga tgcattgtga 21720  
gctcgccctg gtggggagg gctgcaacac tccaggagg ttttgcattt ttcttgcattt ttcttgcattt 21780  
tggttctgca ttctgcattc tacctgatgc cttttatgc ttttgcattt ttcttgcattt ttcttgcattt 21840  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 21900  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 21960  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22020  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22080  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22140  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22200  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22260  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22320  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22380  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22440  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22500  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22560  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22620  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22680  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22740  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22800  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22860  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22920  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 22980  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 23040  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 23100  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 23160  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 23220  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 23280  
ttctgcattc ttttgcattt ttcttgcattt ttcttgcattt ttcttgcattt ccaggttttc 23340

aggggaagtg gggaaaggag ctggggggag ctggggaggtg tgccctgag tgaaggagag 23400  
agggaaaggaa ggaagggtga gactgggcac cttggacttc agtgcagttc taagacatct 23460  
tggcaaggct gatgaggagt tcttgaacca aattcaccag gcaggggagc ctgatgtctc 23520  
aggcaggggc tggcaagtgc agatgcgagg atgttagatt ttggagcaca gcagctggg 23580  
cccttggcta cttccaaggga gctyaggctg gagacctgaa aggcgagttc tcctagctc 23640  
cacaccctt ctccaaggat acaataatat ctgccttata ggattgttgc gagctgagtg 23700  
gcttgcacgtt ctttggaaaga atgaaagcgt atagttatcc caggaagctc aggggttc 23760  
gtgagagctc tggggcttct ccgaagctc ccgagggtgc ttgattcagt tgccagcagg 23820  
gccttccttgc ctggatctt cccccacccc tagccttgc cctcccttc tccttcctt 23880  
cttggaaaggct cagttggccc cccccctccc tccagccacc tggacctgc cagcgtctt 23940  
gtgcaacagg taaagcctac ctgtagcaac aacagatctg ggaaggctc agagggc 24000  
atggggctcg gatcggggc ggctgagacc agagggaaag gtgtgaccct gagtccaccct 24060  
cgctgtcccc gggaaaccac ctcccaggac agctgcctac tgggtctct gccttggaaatt 24120  
gtcacactgc ttttgcacca ggggtccgtt gccccttcc ctttgcgtgg gggaaatgaa 24180  
gttgcgtggag ccgtgagta aactagacct agcagcggg gcacctgtat tggctgtgc 24240  
ctccgggca ggtttcaat gctttcttcc tttttttttcc tggccaggcc acagacggcc 24300  
ctccctttctt gccttgcgtt gtgttctctc agcctcttct gtcttccctt ccaggctggg 24360  
gaataactac atcaactgcgg cgggagccca agtgcgtggc gaggggtctcc gaggcaacac 24420  
ctccctgcag ttcctgggtt aggttggatt ccaggaagag ggacctgtat ggaggggctt 24480  
gggacttttggaggat ttttttttttcc tggccaggcc cagaggcaggc 24540  
ccagctccag tggggaggac aagccaggga gagagtggc ggccttgc tgccacccctt 24600  
atacttggc tatgcctgac aaacagggaa tttggatgt tgggtctagg ggaggacagt 24660  
gcccacacgc tgggtacagg aagccctctg atcctcaggg ggcgtctagg ctgtacttta 24720  
gtgcataattt aaaaaccaccc ggaagcttctt aaacactatt gccaggccctc ccacccca 24780  
ctgatgaat gcaaatatct aggtgcaggccc cccaggatc aggagtttta aaaagcttcc 24840  
caggggatgt acagccagggtt gttggggatcc ctgacccatggg aaagagaagg aaatggggaa 24900  
ggatggaaag gccccaggaa taagagggggc ttttttttttcc tggccaggcc tcttgc 24960  
tgttaggacca tgggttttttttcc tggccaggggag gggagttaccc caacctgcag cccagggtt 25020  
ggtttccctt gtttttttttcc tggccagggtt caccaggctt tgccacccctt ccaggctctct 25080  
cttccctgtt catggcccttgc ttttttttttcc tggccagggtt caccaggcccttcc 25140  
aacctgggac ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt 25200  
aggggggccaa gtttttttttcc tggccagggtt caccaggcccttcc tggccagggtt 25260  
gtctcactt gtttttttttcc tggccagggtt caccaggcccttcc tggccagggtt 25320  
gcccacccatcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt 25380  
cccacccatcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt 25440  
gcccggctgg ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt 25500  
ctgggatttttcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt 25560  
ggactatgtt gtttttttttcc tggccagggtt caccaggcccttcc tggccagggtt 25620  
agaacaaagg agggccacacgc cccactgaac ttttttttttcc tggccagggtt caccaggcccttcc 25680  
tcaggggttat ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt 25740  
gtgttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 25800  
tttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 25860  
tatatttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 25920  
tttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 25980  
gttgcacatca ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26040  
gcccacccatcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26100  
tttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26160  
gatcccttccctt ctttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26220  
ctggcccttcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26280  
tttgcacatca ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26340  
cagggccatcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26400  
agtggacatcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26460  
tcccttccatcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26520  
atgttttttcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26580  
gatggccgttcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26640  
aagaagtttcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26700  
agtttttttcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26760  
cttgcacatcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26820  
gttgcacatcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26880  
tggggccagg ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 26940  
ggtttcccttcc ttttttttttcc tggccagggtt caccaggcccttcc tggccagggtt caccaggcccttcc 27000

acatatacagg tactcactga cactgtctgt tgactctttt ggccctttca gattctgggg 27060  
 caacagagt ggtgacgagg gggcccagc cctggctgaa gccttgggt atcaccagag 27120  
 cttgaggtgg ctcaaggtaa cttcagagtc tatacctgcag tttcttggg gagatcaggt 27180  
 gaagagggag gagctggggc cagttctgaa ggtcttcaa ctttatttct acccccacaat 27240  
 gtttaggcaat ggagtaagga aaaaagacca ttggatttca agagaggaca cttgagtctt 27300  
 tctgggtgac ttgaaatgt cccttgcct ctcagggtt tgatatacgta tctgtaaatt 27360  
 gaagatattt ggctggatca ggtacatttt atcttaaggc ccaattccaa tccattggta 27420  
 gtgggtgccc agtgcaccac attaaaaaga attctaaggc tgcacctggg cttaaagaag 27480  
 agcactataa tcaatttagt atgtctaaa aagctaaaaa aaaaaaaaaa gggcactgca 27540  
 ttcaatttagt gatgtctaaa aagggttagaa aaaaaaaaaa aaagaaaaaa gaaagagcac 27600  
 cgcaatcaat tagtgatgtc taaaatggag cagaccagga gggcaccacg aattttgccc 27660  
 tccataggtt agtcatctc tgaggcttt ccctgcctg acatactttt gttccatgt 27720  
 tacctccagc ctgggggaa acaacattgg cagttgggt gcccaaggc tggactgt 27780  
 gctggcaaag aacgtcatgc tagaagaact ctggtgagtt tgggggattc tctgctctgg 27840  
 ggaagtggat cacaatctct gttgatcccc tggcctcatc cataggagcg gttgtgtgga 27900  
 cagacaaagg tggatgattt agtgatttgc tgatttattt attgtgtttt tctttatata 27960  
 tactgagtgg tatgaagctt atagacccctg gtatgtacat gctaattttt ttatataata 28020  
 aaatatatgg gttgctgtt ttggtgactg cctccacatg gcataagtgt taagagcaca 28080  
 gactctgtaa tcaagcaggc cgtgatctt ggcagttaa ataacaattt cagaatctca 28140  
 agtttcatgt ctgtaaaatg aggtaagaa tacttccaaac cattaaaggat ttttgcaga 28200  
 attagataaa gtagtgcctg tgaagacctt aatatagtgc ctggcatatt tgtaagtgt 28260  
 ccataaatgt taaatttagaa taatggcagg gttactacta ctattactgc tgctgtgt 28320  
 gctgctgtc ctacaactac tatagtactg tgactactac tactaataaa gttttgttat 28380  
 tttaaagtga ttttgagttc cttaggagcac tgggtattca agtcttaggt cattttggaa 28440  
 ggtgtaatgg agttttgata gttgaaagag gaaccatgaa tcatgtttt actgttgacc 28500  
 tgaagcagat tctaaatgg tcatccttta gatgcacta gtatgtttt ctgacatgtt 28560  
 ctggcagct tcagattatg tcagggagat aaaatactga atgtttgatt ttcccgaa 28620  
 gcagaaaggc actgcaacat atgggcattt ccataaaacat attttatgg tggaccttgg 28680  
 ctgttgagg gcttacttgc tctactaaatg tattttttt tctatcctgat ctggattttt 28740  
 ccacttggaa ttctttaga gaggagaacc ttgttatgat agcattttt atgattactg 28800  
 ttaaaagaaa aacttttaggc aaattaaatt tagcagaact ggtttaaca tacagcaatt 28860  
 tatgaattgg gcagcatttca gaactggggat tgctccaccc agcaaggtagt gcaaggatgt 28920  
 tctatagaca ggaaaaggaa gtgtatgtaca aaacagcttgc attgggttgc gctggcatt 28980  
 tgccttatat gggcatggg tttttttt atggatataatg actgtatcagc 29040  
 tggtagactg tgactgactg aagcctggct gctgttattt gctaaagactt agctgttgt 29100  
 tataaggata ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 29160  
 ctcaggccaa atttagtttta actatatgtt aagctgcagg tgacagaata cttccatcta 29220  
 tagaggtttta aacaaggaaa gggtttattt ttccctgtat aggcagctgg atgttaggc 29280  
 ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 29340  
 ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 29400  
 caggatggct gctccagggt cagcactact tctgtattcc cggatttgc ttttttttgc 29460  
 agggaaaggca tctgggttct ctttttttgc ttttttttgc ttttttttgc ttttttttgc 29520  
 tcccttatg tatcaaccat gtgtatgtca ctttttttgc ttttttttgc ttttttttgc 29580  
 cagccctggc, ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 29640  
 aaatggaaaatg agacttccat taataaggaa gaaaggaaatg atggatataatg ggaagctggg 29700  
 ggtttagggaa acttattaca ttggatggcc ttggatgtaa ttggatgttca aatatgtccc 29760  
 tggatgttca aatatgttca ttggatgttca aatatgttca aatatgttca aatatgttca 29820  
 ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 29880  
 ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 29940  
 ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 30000  
 gtaaggaaacc cataaggcagg aaacaggaca ataattgtca gcttttttgc ttttttttgc 30060  
 tgatataatg ctggccgcctt ctttttttgc ttttttttgc ttttttttgc ttttttttgc 30120  
 gcttgccttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 30180  
 ccagttaaatg ctgtataggag agtgggttca gtttttttgc ttttttttgc ttttttttgc 30240  
 ggggagttat caagccaggat atcaatggggat ctttttttgc ttttttttgc ttttttttgc 30300  
 ctggagccag ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 30360  
 cattttatcca ctttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 30420  
 tccagactt ccagctcacc ctttttttgc ttttttttgc ttttttttgc ttttttttgc 30480  
 ctccaggcaga gtttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 30540  
 gatgtatgttgc gtttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 30600  
 tggacagctc aggtgaacat gtttttttgc ttttttttgc ttttttttgc ttttttttgc 30660



aagtctggta aqgccccctgg gcaggccctgt tttagctctc cgaacctca gttttctatc 34380  
 tgtaaaatgg ggtgacggga gagaggaatg gcagaatttt gaggatccct tctgattctg 34440  
 acattcagtg agaatgattc tgcattgtgaa ggatctgatt ctctgtctaa gaaagaagtc 34500  
 tttaccttta taagtaggga gcaatgattt cattttaaa ccttgaactat ttattcagca 34560  
 acttctctgc lctatgyagat agttaggaa tggggatgtg gttgaagaat gaaaagaaaa 34620  
 gtcagctccc gccctccctag aaattgcac tgccttcaca gtcaggat attggatcag 34680  
 accttctgcg gttctgaatg gagattacac aggttaggag cagggtgcac agtgttcca 34740  
 attctctata attaaagcca tagacttca tgtattgaaa aaagcaagaa ttgcatttt 34800  
 gacagattct ttcattgcct taaaaagaat gactgcctt gggagtctgg gcagctgggt 34860  
 ccagtggtgt agactttctc tctgctgagc cacagctca aagatggc tttctgttt 34920  
 ccagggatct atttctcaga caataagtaa aggcttccc tggcctaattg tgctgttaagt 34980  
 gaatgtact atatatgttc caggcactgg gcttagagact aatatttaaa agccagaaaa 35040  
 tttccatata aaaaatctata tctcagggtt ttctcaaaag agctggaaac tctggatgcc 35100  
 cattcatgat tccagtagtt aaccagagta caagaaggc tgagtcttc cagatggca 35160  
 aaccactct ggtgactgc agatccacca agcctattgt cttagaccag gaccctttgg 35220  
 caactcattc ccataagcct gtgacccttg cttaaaatat gcaggccttg tcttctcata 35280  
 aaaagcacat caaggctgca gcgaatgcag atatcaaattg atgaagttaa aaacaaaagc 35340  
 tttgctggc gtggcagctc acacctgtaa tccctagact ttggggagct gaggcaggag 35400  
 gatcaatttta gcccagaggt tcaacaccag accttgc tcaaaaaata aaaaattcag 35460  
 ctgggtgcgg ttagttcct agccactgg gaggctggg tggaaaggatc ccttgaaccc 35520  
 aggagttcaa ggctgctgatg ggcctatgattt gcatcactgc acaggcgcaca gaattagatc 35580  
 ccatctctta aaaaaataaa aaatttaaa gtgacttcaa aatctatgc tgtgtatggag 35640  
 agattttcc ttctgtatga ttgtgatagc tctgtggcct atgacgtcat cagttctgg 35700  
 gcaaagtgtt gttttctgt ttctttgtt ttgaaaccat tgcacagtcc taagaaacat 35760  
 cacattctgg gtccctggca ccagccaaaca tgaggtgagg gcaccagggt ttgctcattt 35820  
 cattcttgc acattcttta attgccttaa aaagaatcac tggccttggg gagtctgtgg 35880  
 ctggctgggt gcagtgttgtt ggactctctc tgcagagtca tggagccttg ttcagaatgc 35940  
 ttccctgagct gccctgggtt gccaagggtt aaaaacagccc tgaacttccct gcaagaaaca 36000  
 ctgcagctgg gccagagagt cagcccatcc caggcatggg tttaaaaagt ggaggcttt 36060  
 gtttgaagc cctgtcttaa ttttgccttc actcaaaccct ctgttcaactt gatctgttt 36120  
 aggctccgag ggaacactttt ctctcttagag gagggtgaca agctcggtg cagggacacc 36180  
 agacttctgc ttgttgcattt cggggaggat gttcgctca gttttttgtt gaggcaggctg 36240  
 tgatgttggg ccccgagggc tgggtgacat gttgtggcag ccttctcaaa atgagccctg 36300  
 tcctgcctaa ggctgaactt gtttctggg aaacccatata gtcaccccttta ttctggcaga 36360  
 ggagggagca tcagtgcctt ccaggataga ctttccaa gcctactttt gccattgtact 36420  
 tcttcccaag attcaatccc agatgtaca aggacagccc ctcctccata gtatggact 36480  
 ggcctctgt gatccctccca ggcttccctg tgggtcagtg gggcccatgg atgtgtttgt 36540  
 taactgagtg ccttttggg gagaggcccc gccttcaca aaagacccctt taccactgt 36600  
 ctgatgaaga ggagttacaca gaacacataa ttcaaggaaagc agctttcccc atgtctcgac 36660  
 tcatccatcc aggcattcc cctgtctgg ttccctccct ctcctggac tcctgcacac 36720  
 gctccctctt ctgaggctga aattcagaat attagtgacc ttagcttga tatttcattt 36780  
 acagcaccccc caaccctggc acccagggtt ggaagggttta caccttagcc tggccctt 36840  
 tccgggtttt aagacattttt tggaaagggtt cacgtgacag ccgtttgttc cccaaagacat 36900  
 tctaggtttt caagaaaaat atgaccacac tccagctggg atcacatgtt gactttattt 36960  
 tccagtgaaa tcagttactc ttcaattttt gctttggaaa cagctgcact tttttttttttt 37020  
 ccaaattgcag cttaaaaaaa ttaatctggg ccagaatttc aaacggcctc actaggctt 37080  
 tggttgc tctgtgactg aactctgaca acagacttct gaaatagacc cacaaggaggc 37140  
 agttccattt catttgcctt agaatgtttt aggatgtaca gttatggatt gaaagtttac 37200  
 agggaaaaaa attaggccgt tccttcaag caaatgtctt cctggattat tcaaaatgtat 37260  
 gtatgttggaa gctttgttattt attgtcagat gctgtgcaaa tgttattttt ttaaaatattt 37320  
 tgatgtgtttaa aaactgggta atattttatg gtcactttgtt ttactgtct taagtttata 37380  
 ctcttataga caacatggcc gtgaacttta tgctgttaat aatcagaggg gaataaaactg 37440  
 ttgttgggtt 37443

<210> 4  
 <211> 1315  
 <212> ADN  
 <213> Homo sapiens

<220>

&lt;221&gt; CDS

&lt;222&gt; (117)..(1118)

&lt;400&gt; 4

cgatcagaag caggtcacac agcctgttc ctgtttcaa acggggact tagaaagtgg 60

cagccctcg gcttgcgccc ggagctgaga accaagagct cgaaggggcc atatga cac 119  
His  
1tcc tcc cgg acc cct gga cac aca cag ccc tgg aga ctg gag cct tgg 167  
Ser Ser Arg Thr Pro Gly His Thr Gln Pro Trp Arg Leu Glu Pro Trp  
5 10 15agc atg gca agt cca gag cac cct ggg agc cct ggc tgc atg gga ccc 215  
Ser Met Ala Ser Pro Glu His Pro Gly Ser Pro Gly Cys Met Gly Pro  
20 25 30ata acc cag tgc acg gca agg acc cag cag gaa gca cca gcc act ggc 263  
Ile Thr Gln Cys Thr Ala Arg Thr Gln Gln Glu Ala Pro Ala Thr Gly  
35 40 45ccc gac ctc ccg cac cca gga cct gac ggg cac tta gac aca cac agt 311  
Pro Asp Leu Pro His Pro Gly Pro Asp Gly His Leu Asp Thr His Ser  
50 55 60 65ggc ctg agc tcc aac tcc agc atg acc acg cgg gag ctt cag cag tac 359  
Gly Leu Ser Ser Asn Ser Met Thr Thr Arg Glu Leu Gln Gln Tyr  
70 75 80tgg cag aac cag aaa tgc cgc tgg aag cac gtc aaa ctg ctc ttt gag 407  
Trp Gln Asn Gln Lys Cys Arg Trp Lys His Val Lys Leu Leu Phe Glu  
85 90 95att gct tca gct cgc atc gag gag aga aaa gtc tct aag ttt gtg gtg 455  
Ile Ala Ser Ala Arg Ile Glu Glu Arg Lys Val Ser Lys Phe Val Val  
100 105 110tac caa atc atc gtc atc cag act ggg agc ttt gac aac aac aag gcc 503  
Tyr Gln Ile Ile Val Ile Gln Thr Gly Ser Phe Asp Asn Asn Lys Ala  
115 120 125gtc ctg gaa cgg cgc tat tcc gac ttc gcg aag ctc cag aaa gcg ctg 551  
Val Leu Glu Arg Arg Tyr Ser Asp Phe Ala Lys Leu Gln Lys Ala Leu  
130 135 140 145ctg aag acg ttc agg gag gag atc gaa gac gtg gag ttt ccc agg aag 599  
Leu Lys Thr Phe Arg Glu Glu Ile Glu Asp Val Glu Phe Pro Arg Lys  
150 155 160cac ctg act ggg aac ttc gct gag gag atc tgt gag cgt cgg cgc 647  
His Leu Thr Gly Asn Phe Ala Glu Glu Met Ile Cys Glu Arg Arg Arg  
165 170 175gcc ctg cag gag tac ctg ggc ctc tac gcc atc cgc tgc gtg cgc 695  
Ala Leu Gln Glu Tyr Leu Gly Leu Leu Tyr Ala Ile Arg Cys Val Arg  
180 185 190cgc tcc cgg gag ttc ctg gac ttc ctc acg cgg ccc gag ctg cgc gag 743  
Arg Ser Arg Glu Phe Leu Asp Phe Leu Thr Arg Pro Glu Leu Arg Glu  
195 200 205

|                                                                     |      |
|---------------------------------------------------------------------|------|
| gct ttc ggc tgc ctg cgg gcc ggc cag tac ccg cgc gcc ctg gag ctg     | 791  |
| Ala Phe Gly Cys Leu Arg Ala Gly Gln Tyr Pro Arg Ala Leu Glu Leu     |      |
| 210 215 220 225                                                     |      |
| ctg ctg cgc gtg ctg ccg ctg cag gag aag ctc acc gcc cac tgc cct     | 839  |
| Leu Leu Arg Val Leu Pro Leu Gln Glu Lys Leu Thr Ala His Cys Pro     |      |
| 230 235 240                                                         |      |
| gcg gcc gcc gtc ccg gcc ctg tgc gcc gtg ctg ctg tgc cac cgc gac     | 887  |
| Ala Ala Ala Val Pro Ala Leu Cys Ala Val Leu Leu Cys His Arg Asp     |      |
| 245 250 255                                                         |      |
| ctc gac cgc ccc gcc gag gcc ttc gcg gcc gga gag agg gcc ctg cag     | 935  |
| Leu Asp Arg Pro Ala Glu Ala Phe Ala Ala Gly Glu Arg Ala Leu Gln     |      |
| 260 265 270                                                         |      |
| cgc ctg cag gcc cgg gag ggc cat cgc tac tat gcg cct ctg ctg gac     | 983  |
| Arg Leu Gln Ala Arg Glu Gly His Arg Tyr Tyr Ala Pro Leu Leu Asp     |      |
| 275 280 285                                                         |      |
| gcc atg gtc cgc ctg gcc tac gcg ctg ggc aag gac ttc gtg act ctg     | 1031 |
| Ala Met Val Arg Leu Ala Tyr Ala Leu Gly Lys Asp Phe Val Thr Leu     |      |
| 290 295 300 305                                                     |      |
| cag gag agg ctg gag gag agc cag ctc cgg agg ccc acg ccc cga ggc     | 1079 |
| Gln Glu Arg Leu Glu Glu Ser Gln Leu Arg Arg Pro Thr Pro Arg Gly     |      |
| 310 315 320                                                         |      |
| atc acc ctg aag gag ctc act gtg cga gaa tac ctg cac tgagccggcc      | 1128 |
| Ile Thr Leu Lys Glu Leu Thr Val Arg Glu Tyr Leu His                 |      |
| 325 330                                                             |      |
| tgggaccccg cagggacgct ggagatgg ggtcaccatg gctcacagtg ggctgttgg      | 1188 |
| ggttttttt tttttttttt cttttttttt tttttttttt gagacagtct tgctctgtca    | 1248 |
| cccaagactga agtgcagtgg ctcaattatg tctcaactgca gcctcaaact cctgggcaca | 1308 |
| agcaatc                                                             | 1315 |

&lt;210&gt; 5

&lt;211&gt; 334

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 5

|                                                                 |  |
|-----------------------------------------------------------------|--|
| His Ser Ser Arg Thr Pro Gly His Thr Gln Pro Trp Arg Leu Glu Pro |  |
| 1 5 10 15                                                       |  |

|                                                                 |  |
|-----------------------------------------------------------------|--|
| Trp Ser Met Ala Ser Pro Glu His Pro Gly Ser Pro Gly Cys Met Gly |  |
| 20 25 30                                                        |  |

|                                                                 |  |
|-----------------------------------------------------------------|--|
| Pro Ile Thr Gln Cys Thr Ala Arg Thr Gln Gln Glu Ala Pro Ala Thr |  |
| 35 40 45                                                        |  |

|                                                                 |  |
|-----------------------------------------------------------------|--|
| Gly Pro Asp Leu Pro His Pro Gly Pro Asp Gly His Leu Asp Thr His |  |
| 50 55 60                                                        |  |

|                                                                 |  |
|-----------------------------------------------------------------|--|
| Ser Gly Leu Ser Ser Asn Ser Ser Met Thr Thr Arg Glu Leu Gln Gln |  |
|-----------------------------------------------------------------|--|

|                                                                 |     |      |     |
|-----------------------------------------------------------------|-----|------|-----|
| 65                                                              | 70  | 75   | 80  |
| Tyr Trp Gln Asn Gln Lys Cys Arg Trp Lys His Val Leu Leu Phe     |     |      |     |
| 85                                                              | 90  | 95   |     |
| Glu Ile Ala Ser Ala Arg Ile Glu Glu Arg Lys Val Ser Lys Phe Val |     |      |     |
| 100                                                             | 105 | 110  |     |
| Val Tyr Gln Ile Ile Val Ile Gln Thr Gly Ser Phe Asp Asn Asn Lys |     |      |     |
| 115                                                             | 120 | 125  |     |
| Ala Val Leu Glu Arg Arg Tyr Ser Asp Phe Ala Lys Leu Gln Lys Ala |     |      |     |
| 130                                                             | 135 | 140  |     |
| Leu Leu Lys Thr Phe Arg Glu Glu Ile Glu Asp Val Glu Phe Pro Arg |     |      |     |
| 145                                                             | 150 | 155  | 160 |
| Lys His Leu Thr Gly Asn Phe Ala Glu Glu Met Ile Cys Glu Arg Arg |     |      |     |
| 165                                                             | 170 | 175  |     |
| Arg Ala Leu Gln Glu Tyr Leu Gly Leu Leu Tyr Ala Ile Arg Cys Val |     |      |     |
| 180                                                             | 185 | 190  |     |
| Arg Arg Ser Arg Glu Phe Leu Asp Phe Leu Thr Arg Pro Glu Leu Arg |     |      |     |
| 195                                                             | 200 | 205  |     |
| Glu Ala Phe Gly Cys Leu Arg Ala Gly Gln Tyr Pro Arg Ala Leu Glu |     |      |     |
| 210                                                             | 215 | 220  |     |
| Leu Leu Leu Arg Val Leu Pro Leu Gln Glu Lys Leu Thr Ala His Cys |     |      |     |
| 225                                                             | 230 | 235  | 240 |
| Pro Ala Ala Ala Val Pro Ala Leu Cys Ala Val Leu Leu Cys His Arg |     |      |     |
| 245                                                             | 250 | 255  |     |
| Asp Leu Asp Arg Pro Ala Glu Ala Phe Ala Ala Gly Glu Arg Ala Leu |     |      |     |
| 260                                                             | 265 | 270  |     |
| Gln Arg Leu Gln Ala Arg Glu Gly His Arg Tyr Tyr Ala Pro Leu Leu |     |      |     |
| 275                                                             | 280 | 285  |     |
| Asp Ala Met Val Arg Leu Ala Tyr Ala Leu Gly Lys Asp Phe Val Thr |     |      |     |
| 290                                                             | 295 | 300. |     |
| Leu Gln Glu Arg Leu Glu Glu Ser Gln Leu Arg Arg Pro Thr Pro Arg |     |      |     |
| 305                                                             | 310 | 315  | 320 |
| Gly Ile Thr Leu Lys Glu Leu Thr Val Arg Glu Tyr Leu His         |     |      |     |
| 325                                                             | 330 |      |     |

<210> 6  
 <211> 8135  
 <212> ADN  
 <213> Homo sapiens

<220>  
 <221> exon  
 <222> (1)..(161)

<220>  
<221> exon  
<222> (3812)..(3950)

<220>  
<221> exon  
<222> (5426)..(5577)

<220>  
<221> exon  
<222> (7273)..(8135)

<400> 6  
cgatcagaag caggtcacac agcctgttc ctgtttcaa acggggact tagaaagtgg 60  
cagccctcg gcttgtcgcc ggagctgaga accaagagct cgaaggggccc atatgacact 120  
cctcccgac ccctggacac acacagccct ggagactgga ggtcagtatt tgatccaaag 180  
ctcagctgtc ctctgcctgc tggcgttgc gttcccttct cctggggccc tgcctggcac 240  
ctgctgggg cagggtggga gggggaaagag ttagtgacag ccgctgtgtc tggagctctc 300  
cttagcacac tgagggcagag gaagggacag ctccctggacc ttccatcacc tccattcctt 360  
ttgaaatgtc aggccgttgc acaacccatc ttggccttgc agaataagtc accacaccc 420  
tgtttctaa aagaacagtg tcagggaaacc cctgcctcag cacagcctta gaggactcat 480  
ggaaaatgca gaatccaggc ctgttcaatg gcaccccttct atgttagcag ccagggaaacc 540  
tgctcttggc caagccccctg ggatccccc cccacccccc caggggattt ttacacacac 600  
tgggttggga gcccctggct ttggcaaggc ttctcaggtg agcgtccagt tggaggg 660  
tacccaccct ttccccaaga gaggcagcca cacatccaac atccctggat ctctgtctcc 720  
cagcgtggc catgtgttccatttccatc ctagaggcctc atccccccatg aaaagtcttc 780  
cgccaggccct cagaagata gtgtggccctc tggcgttgc cagaagaag gactggactt 840  
ggcagtcgc tcttggagag ggggtgggta ggacacccctgg ggacaggagg aggagaatga 900  
ctgtctgtc acacacggct ggaaggtaaca ggaggctggg aagctgtctc gtccctgggg 960  
ccaactacag gccccccaggc caacagcaac aacacttttta gtatttgtt ataaagtcaa 1020  
gaaatcttttgc ttagcagaggc tgaggagagg gaagggaaagg gcacatggaaac cgtctatgt 1080  
gctatccccca gagagctttt agagtgacag gattgtttc ccatttcaca gatgaggaaa 1140  
ctgaggcctg gagagggtatg ggaagctacc caaggccccca tggatcaccc agtgcacaac 1200  
tctttcttc cccctcccttct ttaaatgggt gattcccaat gaaacctgtt agagacaacc 1260  
ataaggggc tgactgtggc tgctgaatg gattttttaatc tggggcttggg ttttataatc 1320  
agctttctca gtcttactg gagggtcaag ccggggatccatttctggg tcttacagg 1380  
tctctggcc aatagtgcctc tgcttctgc ctggggccat ctgcctggc atgaaagcag 1440  
atctgaaag gctggggccc ctggggccat ggcacttcgc catccccat tttacagaag 1500  
tgctgagcat aggagtgcctc tggggccccca agaattcccg ccaccaagaa tcacgtaaac 1560  
catccactgt ctcaacttggc caccatgc aatgttaggg acccaccctt agtcatccat 1620  
catcttatca acaggacggg gctttagtgc acatttata ggttagggaaa ctgaagccca 1680  
gagatattaa agcaacttgc taaggacaca cgggttgtca ggatggaaagg cgatgtctcc 1740  
tgactccctg acaggcacaa gagacaagcg agaggtggcc gtagccat gctcaagaac 1800  
gtgcagccct gggccagcca ggccccctgtt ccgtgcctct gtttgcctt ctgtaaaagg 1860  
tgaggttggc tggggggccat ctggggccat cccactggat ggctgtcag agccaaacgg 1920  
agaaggcccccc aggggttccctt tcacccgaca cagcaagcac ttcccccgtt agtgcaggt 1980  
ccaggccccca gctgacccctt cctctccctg gccagccgtt ctcacccctt gggcaaggaa 2040  
caggcgctgg ctgtgtcag ggacatgc gactccggcc cccatctgtt ctcagggggt 2100  
gccaggggagg cactggctct atcttcttctt agggccgtt cagcccaggg gttcagacca 2160  
agagccccaga atccaacaga tcagaggatc agtcccgatcttacccat gttccactgg 2220  
cagcttccctt aggtcatttgc cacccttgcattt gtttgcatttccatgcctt accagtataac 2280  
cagctactcc ctccagccga tctaattttttaatttgcattt ttttgcatttcaatg ttgtctcaaa 2340  
catttgcattt tctattccaa tccacccat ttttgcattt tatttgcatttcaatg atatttctgg 2400  
aaacatcttag cacttaacag acactaaaag cgggggtact acacagtccc tgggatggac 2460  
agggccctga gctgaggctt cagagtctgc ctgactgtat cctcacccttca gccttgcggaa 2520  
cgtgggttctt gtttattatcc ccaattttata gggaaacagaa gcacagagaa gttgagtcac 2580  
ttggccatca ccagggtcattt ccttccactt atccgggtca cagacagagt tattatgtaa 2640  
accagatccc agtgcctgt tctccctccctt tgtagttaggt ggagagaatt ctgaagtcac 2700  
cccagccctgg gctgtatcc tggccaccac tccaccatgc ctcacccat tggccatca 2760  
gtctcgttcc ccttacatca aaaggggatg gtaaacagtc ctgactgtcag acagtgttca 2820  
gttttagtgc agagtgtgtc ctgggtgtca agtgcacacgc cagcactgtca caagcactgg 2880

agacaaatc agcttgctt gttgcgcaca ctcaccagct gcggtacttt agacccatgt 2940  
tttctcatct gttatgtggt ggtaatgata gactttgtg agcattaaac tagatttaggg 3000  
gctatggaga acctagatgg gtatgaagt ggtataataa gctatcggt aattttctgt 3060  
atagatagat tattgattga ttgatcgata gaagattcat accagtatct acctgctctg 3120  
aacactgacc tttcttttt tcttttttag ggtataataa gctatcggt aattttctgt 3060  
gcagtgccat catcatagct cactgcagcc atggtctgt tctgtcaccc agactggagt 3180  
tctcagccct ccaagtagct gggaccacag tcagtcctt gggcttaagg gatcctctg 3240  
tttcttagaga cggggctca ctacattggc gcgtgcattcc tggataattt ttttttattt 3300  
gatccttca acccagcctc ccaaagcgct caggctgtc tcaaatttcc gggctcaagt 3360  
cttgaacact gagacttcat tcgcattgtt gggattacag gcatgatgg ccatgttcaa 3420  
catcttttca tcaagtaatc actaaagcca aacataaaac tgagttatcta gacaagccag 3480  
actctgagca atacgtaagg atcacctcaa atactttac ttgaatcat ctcatttcaa 3540  
gtcttctctg ccttggagta acctgcccag taacatatgg atcatgc当地 taggtgaagg 3600  
cagctggag agtggggaaag gttgagccgt caaaggggca gaccagatt tggatctgg 3660  
gagggggcat gacacagctc ctaggcaccc gggcccttgc tcatccctc tgctgcccag 3720  
ggtcctact gttctttt tecttgcggc caggagccac cgggaacccc aactggagtg 3780  
tggggccct ggctgcattgg gaccataaac gccttggagc atggcaagtc cagagcaccc 3840  
accagccat ggcccccggc tccgcaccc ccagtgcacg gcaaggagcc agcaggaagc 3900  
ggcttgcagac tcggcttggg ggaggggtt ggacacttag tggtggctga 3960  
tttggggat gcagcaagag gcccggcag gaagacatc aaagttacaaa tggtggctac 4020  
cactgagaca cagatctgtt gcaagctgtt ctttgttaac ttgggttatac cccaaaacaga 4080  
agggggatgg gaatggaagg aaagggcaag gggccaggcc aggacatcag tgaacagata 4200  
ggcacggtag gtggctgaag ctcacccca gcttgcatttgc tcttcgcagat 4260  
atctctgggt tgccttatcc taggggtgag gaagccggcgc tgttatctac cagtctgc 4320  
ctgcataatggga agggggacgc tccctggccct gctgtatgg ccctagaaaag ccctcaggga 4380  
agccagtggc atgtttcttga aaagtgggtt ccaaggggc acgggtccacg ctggggcatg 4440  
gacagcatct gctgtatgtc catctccttgc aacagatctt ttcttacagt ctttcgagat 4500  
gcccatttca atacctgctc tggttgc gctatgcagg gcactggaga aacagaaaaca 4560  
ggaagaaatc aaacactgca ctgttgc ggtttggtag agaaaacagat cagtggaaaa 4620  
cagttacacg tgccacgaga aataaataaaa taaaatgaaa aacctgttagg aacaagggtgg 4680  
gaagcttta ctctaatgccc aaggggcatt tgcaagtatgg tggtggctgg gtcttgc 4740  
gttagacttgg aaagggtctgg gaccatgttgc ctttgcataa aaatgcacaa ttatttgc 4800  
ttcttaagaa cctcagatgtt ggcagggtt caagtgggtt ttaagaaaca ctgtgttgc 4860  
tttccaggcg tggaaataga ggggtggatg caaggcagag cagtgcacgt ccgagaagag 4920  
cccgcatgtt gggcgttag atgagaaggt taggaaggcc cagcccgtg aggctggaaac 4980  
ataacatctt cctcaactgccc tccctgc gctatgtgtt gctcaaggag tcgtggcaac 5040  
agtacacgaa tcagggttgc agggagcaca gaaacacaca agccaccgtc tctgttgc 5100  
cagagcagggtt atttcaccat ggcacatctt cagaccagaa gtggacgtatg caaagtgc 5160  
gcaccgcattt ccaaagctgtt gaaaccactt ggggggtatg ggcttatttgg gattgtcggt 5220  
ggtaggggtt attctgcac gctgggcaca gagggtcttc tgatgccccca attgggccta 5280  
taaatggcg ggtggggagag agggatattt aatacttttc aggagttctg atatgcacatc 5340  
tcagatagac ccagccatctt ccccaagccc atgcctcgga agtgcactga cagggtgcag 5400  
atcccttaagg gtgttgcctt tccagacaca cacagtggcc tgatgtccaa ctccagcatg 5460  
accacgcggg agtctcagca gtactggcag aaccagaaat gccgtggaa gcacgtcaaa 5520  
ctgtcttttgc agatcgctt agtgcgcattt gaggagggaa aagtcttcaa ttgttgc 5580  
agcagagatgtt gggaaatgtt ggagcctttt tcactctgtc ttcttgc tcttgcataaa 5640  
gtcttgcataa gcttcaggtt tcccaactt gaaatgggtc aacacactaa ctacacgtt 5700  
tcttgcataa aaaatggccca aagagcaaga tttcagggtc agcacctgtt aggggtgtt 5760  
aggattcgaa ccatataatgtt catattttt ggtcccaaga agggaaatagc ccagttaat 5820  
cccatcttcatc cagggttgc tccatgttgc cttttcttca ccaatttgc catabactg 5880  
tatctgttctt aatttattt atttattttt ttctttaat tggatcactt tttaaaaaca 5940  
tgaagcacat ttatccaatg gggaaatacc ttaaatggaa aaccaatatac acatggcaca 6000  
aagcaaaatg aacataacttag aaaatgtcgat acaaggaaag tcaatacaag gaaaagctatg 6060  
tgctgttattt aaattcttagc tggttactgtt ggcttgcggaa aagccctgtt cttggagct 6120  
gtctcttccttcc ctgtttagat gggattttag cttgtttaa gggatgtttaa agatggctta 6180  
agagccacac ttcatcttc tccatgttgc acctggacc gggataataa acatagctac 6240  
caactgatgc caatggcatg cccggcacag ctccatgttgc ttgcgttgc ttaacttcat 6300  
taatccctac tgggtgaggtt aggcactatg cctatcttgc ttatgttgc gggatgtttaa 6360  
agactcgag aggtttaaattt actcatctaa aaccacacag cttagaccatg gttagggctat 6420  
aattacaacc catcaatctt ggcttgcggag tcaatgttgc gggatgttataa tgcccttaat 6480  
atataatttc cctgtatcatc gattcttgc aaagatgtt gaaaaggattt gatgttcttca 6540

ccatataacg gcatcaccag tgtacctaata tgatgttata ttgtacgtaa aactaattcc 6600  
 caagtgtgaa acatttgaa aacacagcat ctcagttcag aaaacagagg cccagttta 6660  
 gcaagtaaag ccaagaggaa cccacgcgc ctgcaggcga ggaccctctg cccttcctcc 6720  
 tcccaatgtt ccccaccttgc tggtgttgc ttccagggt tgactcagct gatgccaata 6780  
 gcaattaaa acagaattgg gccagggtc glggctcatg cclytaatcc cagcacttg 6840  
 ggaggcccggat gtggggatcg cgcttgagcc caggagttgg agaccgcctt gggcaacaca 6900  
 gccagacccc atctttaaa aagaatcaa aaatctgcca ggttagtgggt gtgcctgttag 6960  
 tcccaatgtt ccaagggtc cagggtggca ggtcaatttg gcccataatgt tcaagggttc 7020  
 agtgggtat gatcgcata cttgtacttgc gctgtggtaa cagtgcgaga ccctgtctc 7080  
 aaaaataat aaataaataa ataaataat aaataaaacaa acaaacaacaa aacaaacaa 7140  
 tcaattgtcat ataaggatcg cccgtttca gggcatgctt tacaccggcc tggtaactt 7200  
 tactctgggt gtgcgtccgtc cggcgacgc cccggccggaa ggtggccaca gctctctcg 7260  
 gttgcgcctt aggtgtacca aatcatgtc atccagactg ggagcttga caacaacaag 7320  
 gccgtcctgg aacggcgcta ttccgacttc gcaagactcc agaaaggcgt gctgaagacg 7380  
 ttcaaggagg agatcgaaga cgtggagtt cccaggaagc acctgactgg gaaacttcgct 7440  
 gaggagatga tctgtgagcg tggcgccgc ctgcaggagt acctggccct gctctacgccc 7500  
 atccgctqcg tgcgcccgtc cggggagttc ctggacttcc tcaacggccgc ggagctgcgc 7560  
 gaggcttcg gtcgctgcg ggccggccag taccggcgccg ccctggagct gctgctgcgc 7620  
 gtgcgtccgc tgcaggagaa gtcacccgc cactgcccgt cggccggccgt cccggccctg 7680  
 tgcgcccgtc tgctgtgcca cggcgaccc tggccggccgtt cggggccggaa 7740  
 gagaggggccc tgcagcgcct gcaggcccgg gaggggccatc getactatgc gcctctgtc 7800  
 gacgcccatttgg tccgcctggc ctacgcgtc ggcaaggact tcgtgactct gcaggagagg 7860  
 ctggaggaga gccagctccg gaggcccacg ccccgaggca tcaaccctgaa ggagctcact 7920  
 gtgcgagaat acctgcactg agccggccgt gggacccgc gggacgcctt agatttgggg 7980  
 tcaccatggc tcacagtggg ctgtttgggg ttcttttttt ttattttcc ttttctttt 8040  
 tggtaatttga gacagtcttgc ctctgtcacc cagactgaag tgcagtggtt caattatgtc 8100  
 tcactgcagc ctcaaactcc tgggcacaag caatc 8135

<210> 7  
 <211> 16  
 <212> ADN  
 <213> Homo sapiens

<400> 7  
 ctgggtgcga ttgtc

16

<210> 8  
 <211> 16  
 <212> ADN  
 <213> Homo sapiens

<400> 8  
 ccaggccccca tgacag

16

<210> 9  
 <211> 25  
 <212> ADN  
 <213> Homo sapiens

<400> 9  
 tggtccggc ccaatccaa tgctt

25

<210> 10  
 <211> 28  
 <212> ADN  
 <213> Homo sapiens

|                                |    |
|--------------------------------|----|
| <400> 10                       |    |
| ttcctcatgt ataaattggg tgtggcca | 28 |
| <210> 11                       |    |
| <211> 25                       |    |
| <212> ADN                      |    |
| <213> Homo sapiens             |    |
| <400> 11                       |    |
| acagagttag gaccccatct ctatc    | 25 |
| <210> 12                       |    |
| <211> 25                       |    |
| <212> ADN                      |    |
| <213> Homo sapiens             |    |
| <400> 12                       |    |
| tccaaactgct gggattacag gcaca   | 25 |
| <210> 13                       |    |
| <211> 22                       |    |
| <212> ADN                      |    |
| <213> Homo sapiens             |    |
| <400> 13                       |    |
| agtcccccag accagggcaa ac       | 22 |
| <210> 14                       |    |
| <211> 23                       |    |
| <212> ADN                      |    |
| <213> Homo sapiens             |    |
| <400> 14                       |    |
| tccatttctg cagtacacat gca      | 23 |
| <210> 15                       |    |
| <211> 20                       |    |
| <212> ADN                      |    |
| <213> Homo sapiens             |    |
| <400> 15                       |    |
| ctctccccat agaaggcata          | 20 |
| <210> 16                       |    |
| <211> 20                       |    |
| <212> ADN                      |    |
| <213> Homo sapiens             |    |
| <400> 16                       |    |
| ggatagagac gttctcttaa          | 20 |
| <210> 17                       |    |
| <211> 20                       |    |
| <212> ADN                      |    |

|                             |    |
|-----------------------------|----|
| <213> Homo sapiens          |    |
| <400> 17                    |    |
| caggctgaat gacagaacaa       | 20 |
| <210> 18                    |    |
| <211> 20                    |    |
| <212> ADN                   |    |
| <213> Homo sapiens          |    |
| <400> 18                    |    |
| attgaaaaca actccgtcca       | 20 |
| <210> 19                    |    |
| <211> 25                    |    |
| <212> ADN                   |    |
| <213> Homo sapiens          |    |
| <400> 19                    |    |
| atactcaatt ttagacagtt caggg | 25 |
| <210> 20                    |    |
| <211> 21                    |    |
| <212> ADN                   |    |
| <213> Homo sapiens          |    |
| <400> 20                    |    |
| ggctcagttc ctaaccagtt c     | 21 |
| <210> 21                    |    |
| <211> 20                    |    |
| <212> ADN                   |    |
| <213> Homo sapiens          |    |
| <400> 21                    |    |
| agtcaagtctg tccagaggtg      | 20 |
| <210> 22                    |    |
| <211> 20                    |    |
| <212> ADN                   |    |
| <213> Homo sapiens          |    |
| <400> 22                    |    |
| tgaatcttac atccccatccc      | 20 |
| <210> 23                    |    |
| <211> 17                    |    |
| <212> ADN                   |    |
| <213> Homo sapiens          |    |
| <400> 23                    |    |
| gatcttccca aagcgcc          | 17 |
| <210> 24                    |    |

<211> 17  
<212> ADN  
<213> Homo sapiens

<400> 24  
tcccgctcagc caagcta

17

<210> 25  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 25  
aagcttgtat ctttctcagg

20

<210> 26  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 26  
atctacacctg gctgtcattg

20

<210> 27  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 27  
cctccataat catgtgagcc

20

<210> 28  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 28  
aatctcccca actcaagacc

20

<210> 29  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 29  
ggatgcctgc tctaaatacc

20

<210> 30  
<211> 19  
<212> ADN  
<213> Homo sapiens

<400> 30  
cccaggggtc aaacttaat

19

<210> 31  
<211> 21  
<212> ADN  
<213> Homo sapiens

<400> 31  
ggtttcaaag tatctccagg g

21

<210> 32  
<211> 21  
<212> ADN  
<213> Homo sapiens

<400> 32  
ggtttcaaag tatctccagg g

21

<210> 33  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 33  
gtgcgtgtgt tcgttatcaac

20

<210> 34  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 34  
tcatctccaa aggagttct

20

<210> 35  
<211> 18  
<212> ADN  
<213> Homo sapiens

<400> 35  
aaagcccaacc ttgcttca

18

<210> 36  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 36  
tcttgaaac aggttaagtgc

20

<210> 37  
<211> 18  
<212> ADN  
<213> Homo sapiens

<400> 37

attgccctca agaacagc

18

<210> 38  
 <211> 17  
 <212> ADN  
 <213> Homo sapiens

<400> 38  
 gtgctatgcc atccag

17

<210> 39  
 <211> 20  
 <212> ADN  
 <213> Homo sapiens

<400> 39  
 ccacaccagc gttttctaa

20

<210> 40  
 <211> 24  
 <212> ADN  
 <213> Homo sapiens

<400> 40  
 cacacttac acacacctat accc

24

<210> 41  
 <211> 22  
 <212> ADN  
 <213> Homo sapiens

<400> 41  
 aagccatatt aggtctgtcc at

22

<210> 42  
 <211> 19  
 <212> ADN  
 <213> Homo sapiens

<400> 42  
 gcttgggtta aatgcgtgt

19

<210> 43  
 <211> 20  
 <212> ADN  
 <213> Homo sapiens

<400> 43  
 agcagtttgg gtaaacattg

20

<210> 44  
 <211> 20  
 <212> ADN  
 <213> Homo sapiens

|                             |    |
|-----------------------------|----|
| <400> 44                    |    |
| aaatatgcct tctggagggtg      | 20 |
|                             |    |
| <210> 45                    |    |
| <211> 20                    |    |
| <212> ADN                   |    |
| <213> Homo sapiens          |    |
|                             |    |
| <400> 45                    |    |
| ggaggatcag gggagtttat       | 20 |
|                             |    |
| <210> 46                    |    |
| <211> 24                    |    |
| <212> ADN                   |    |
| <213> Homo sapiens          |    |
|                             |    |
| <400> 46                    |    |
| caaagtaaat gaatgtctac tgcc  | 24 |
|                             |    |
| <210> 47                    |    |
| <211> 23                    |    |
| <212> ADN                   |    |
| <213> Homo sapiens          |    |
|                             |    |
| <400> 47                    |    |
| ccaaactctgt agtttcaaag agc  | 23 |
|                             |    |
| <210> 48                    |    |
| <211> 20                    |    |
| <212> ADN                   |    |
| <213> Homo sapiens          |    |
|                             |    |
| <400> 48                    |    |
| tcacagccta cttgcttggt       | 20 |
|                             |    |
| <210> 49                    |    |
| <211> 25                    |    |
| <212> ADN                   |    |
| <213> Homo sapiens          |    |
|                             |    |
| <400> 49                    |    |
| gacagcctca aatgaaatat aacac | 25 |
|                             |    |
| <210> 50                    |    |
| <211> 25                    |    |
| <212> ADN                   |    |
| <213> Homo sapiens          |    |
|                             |    |
| <400> 50                    |    |
| gctctcagct agggtagttg tttat | 25 |
|                             |    |
| <210> 51                    |    |
| <211> 25                    |    |

|                             |  |    |
|-----------------------------|--|----|
| <212> ADN                   |  |    |
| <213> Homo sapiens          |  |    |
| <400> 51                    |  |    |
| atttttaagg aatgtaaagn acaca |  | 25 |
|                             |  |    |
| <210> 52                    |  |    |
| <211> 20                    |  |    |
| <212> ADN                   |  |    |
| <213> Homo sapiens          |  |    |
| <400> 52                    |  |    |
| gaccaggagt cagtaaaagg       |  | 20 |
|                             |  |    |
| <210> 53                    |  |    |
| <211> 20                    |  |    |
| <212> ADN                   |  |    |
| <213> Homo sapiens          |  |    |
| <400> 53                    |  |    |
| gtccaaaaca ccaccctcta       |  | 20 |
|                             |  |    |
| <210> 54                    |  |    |
| <211> 24                    |  |    |
| <212> ADN                   |  |    |
| <213> Homo sapiens          |  |    |
| <400> 54                    |  |    |
| gaagtagatc agtcatcttg ctgc  |  | 24 |
|                             |  |    |
| <210> 55                    |  |    |
| <211> 19                    |  |    |
| <212> ADN                   |  |    |
| <213> Homo sapiens          |  |    |
| <400> 55                    |  |    |
| tcctctgggg gattcactc        |  | 19 |
|                             |  |    |
| <210> 56                    |  |    |
| <211> 20                    |  |    |
| <212> ADN                   |  |    |
| <213> Homo sapiens          |  |    |
| <400> 56                    |  |    |
| gggacatcac caagcacaag       |  | 20 |
|                             |  |    |
| <210> 57                    |  |    |
| <211> 25                    |  |    |
| <212> ADN                   |  |    |
| <213> Homo sapiens          |  |    |
| <400> 57                    |  |    |
| caggaaaata aatctaacac acata |  | 25 |

<210> 58  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 58  
cctgtggca ctgataaata

20

<210> 59  
<211> 19  
<212> ADN  
<213> Homo sapiens

<400> 59  
cccagccccc atctcacccg

19

<210> 60  
<211> 19  
<212> ADN  
<213> Homo sapiens

<400> 60  
cccagccccc atctcacca

19

<210> 61  
<211> 19  
<212> ADN  
<213> Homo sapiens

<400> 61  
ctgcggagga ggctgctgg

19

<210> 62  
<211> 19  
<212> ADN  
<213> Homo sapiens

<400> 62  
tcactcccac cacccttcc

19

<210> 63  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 63  
agaagtttag tgtggcgtgg

20

<210> 64  
<211> 17  
<212> ADN  
<213> Homo sapiens

<400> 64  
gccatctccc caagccc

17

|                       |    |
|-----------------------|----|
| <210> 65              |    |
| <211> 18              |    |
| <212> ADN             |    |
| <213> Homo sapiens    |    |
| <br>                  |    |
| <400> 65              |    |
| tcgatgcgag ctgaagcg   | 18 |
| <br>                  |    |
| <210> 66              |    |
| <211> 18              |    |
| <212> ADN             |    |
| <213> Homo sapiens    |    |
| <br>                  |    |
| <400> 66              |    |
| tcgatgcgag ctgaagca   | 18 |
| <br>                  |    |
| <210> 67              |    |
| <211> 20              |    |
| <212> ADN             |    |
| <213> Homo sapiens    |    |
| <br>                  |    |
| <400> 67              |    |
| tgaatgttaa agggctctgg | 20 |
| <br>                  |    |
| <210> 68              |    |
| <211> 19              |    |
| <212> ADN             |    |
| <213> Homo sapiens    |    |
| <br>                  |    |
| <400> 68              |    |
| ttggttctca gctccggcg  | 19 |
| <br>                  |    |
| <210> 69              |    |
| <211> 19              |    |
| <212> ADN             |    |
| <213> Homo sapiens    |    |
| <br>                  |    |
| <400> 69              |    |
| ttggttctca gctccggca  | 19 |
| <br>                  |    |
| <210> 70              |    |
| <211> 19              |    |
| <212> ADN             |    |
| <213> Homo sapiens    |    |
| <br>                  |    |
| <400> 70              |    |
| agaaaaccggg ctggctgtg | 19 |
| <br>                  |    |
| <210> 71              |    |
| <211> 21              |    |
| <212> ADN             |    |
| <213> Homo sapiens    |    |

<400> 71  
gcattgcctt ttgatctcta c

21

<210> 72  
<211> 18  
<212> ADN  
<213> Homo sapiens

<400> 72  
tgggctcttc tgccccca

18

<210> 73  
<211> 18  
<212> ADN  
<213> Homo sapiens

<400> 73  
tgggctcttc tgccccca

18

<210> 74  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 74  
tgcctttct tctgccttcc

20

<210> 75  
<211> 22  
<212> ADN  
<213> Homo sapiens

<400> 75  
cgagctgtac ctgaggaagc gt

22

<210> 76  
<211> 24  
<212> ADN  
<213> Homo sapiens

<400> 76  
cctgagctgt acctgagaa gcgc

24

<210> 77  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 77  
catcatgagc ccgggggtggc

20

<210> 78  
<211> 23  
<212> ADN

&lt;213&gt; Homo sapiens

&lt;400&gt; 78

tttctttgg ctccctggtg cgt

23

&lt;210&gt; 79

&lt;211&gt; 25

&lt;212&gt; ADN

&lt;213&gt; Homo sapiens

&lt;400&gt; 79

accttctctt ggcttcctgg tgcgg

25

&lt;210&gt; 80

&lt;211&gt; 26

&lt;212&gt; ADN

&lt;213&gt; Homo sapiens

&lt;400&gt; 80

gccaaaggta tcgtgccagg gctcca

26

&lt;210&gt; 81

&lt;211&gt; 20

&lt;212&gt; ADN

&lt;213&gt; Homo sapiens

&lt;400&gt; 81

atctgagaag gccctgctct

20

&lt;210&gt; 82

&lt;211&gt; 20

&lt;212&gt; ADN

&lt;213&gt; Homo sapiens

&lt;400&gt; 82

atctgagaag gccctgctcc

20

&lt;210&gt; 83

&lt;211&gt; 19

&lt;212&gt; ADN

&lt;213&gt; Homo sapiens

&lt;400&gt; 83

cccacactta gccttgatg

19

&lt;210&gt; 84

&lt;211&gt; 19

&lt;212&gt; ADN

&lt;213&gt; Homo sapiens

&lt;400&gt; 84

atgagttgc ccagcggag

19

&lt;210&gt; 85

WO 01/72822

PCT/FR01/00935

<211> 19  
<212> ADN  
<213> Homo sapiens

<400> 85  
attgagagcc cttggagtg

19

<210> 86  
<211> 19  
<212> ADN  
<213> Homo sapiens

<400> 86  
tgatttcgta agacaagtg

19

<210> 87  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 87  
agcaaattctt aggagttatg

20

<210> 88  
<211> 19  
<212> ADN  
<213> Homo sapiens

<400> 88  
agctgagatg tccggatcg

19

<210> 89  
<211> 18  
<212> ADN  
<213> Homo sapiens

<400> 89  
agctgagattt ccggatca

18

<210> 90  
<211> 20  
<212> ADN  
<213> Homo sapiens

<400> 90  
gtcctcttaa cttcccttcc

20